

Carolyn Müller-Spitzer (Mannheim)

Methoden der Wörterbuchbenutzungsforschung¹

1	Zum Stand der Wörterbuchbenutzungsforschung	3.2.1	Einleitung
2	Methodisch-terminologische Grundlagen	3.3.2	Methode
3	Das methodische Spektrum innerhalb der Benutzungsforschung verdeutlicht anhand einzelner Studien	3.2.3	Ergebnisse und Diskussion
		3.3	Auffällig oft nachgeschlagene Stichwörter im deutschen Wiktionary
3.1	Bildschirmgestaltung von Onlinewörterbüchern	3.3.1	Einleitung
3.1.1	Einleitung	3.3.2	Methode
3.1.2	Methode	3.3.3	Ergebnis und Diskussion
3.1.3	Ergebnisse und Diskussion	3.4	Wie navigieren potenzielle Nutzer zu einzelnen Bedeutungen im Onlinewörterbuch eleXiko?
3.2	Die Einschätzung innovativer Features von Onlinewörterbüchern	4	Schlussbemerkung
		5	Bibliographie

Abstract: The methods utilized in the area of research into dictionary use are established research methods in the social sciences. After explicating the different steps of a typical empirical investigation, this article provides examples of how these different methods are used in various user studies conducted in the field of using online dictionaries. Thereby, different kinds of data collection (surveys as online questionnaires, log files and eye tracking) as well as different research design structures (for instance, ex-post-facto design or experimental design) are discussed.

Keywords: empirische Wörterbuchbenutzungsforschung, Onlinewörterbuch, Befragung, Eyetracking, Logfiles

¹ Einen ähnlichen Überblick zur Wörterbuchbenutzungsforschung – allerdings mehr auf Studierende zugeschnitten – bietet der Beitrag „Benutzungsforschung“ von mir im „Studienbuch Internetlexikografie“, welches vom DFG-Netzwerk „Internetlexikografie“ herausgegeben wird und 2015 bei de Gruyter erscheinen soll. Sämtliche Studien, die hier erwähnt werden, sind außerdem ausführlich dokumentiert in Müller-Spitzer 2014a.

1 Zum Stand der Wörterbuchbenutzungsforschung

Die Wörterbuchbenutzungsforschung gilt als jüngster Forschungszweig innerhalb der Wörterbuchforschung (Wiegand 1998, 259). Dass die Bedeutung dieses Forschungszweiges in den letzten Jahren zugenommen hat, ist das Verdienst vieler Lexikografinnen und Wörterbuchforscher (s. Hinweise dazu bei Rundell 2012b, 3; Tarp 2009, 276). Zwar wurde in einzelnen Publikationen schon lange betont, dass die Benutzer eine zentrale Größe bei der Konzeptionierung lexikografischer Prozesse sein sollen (cf. e.g. Householder 1962; Wiegand 1977); mittlerweile kann es aber – anders als noch vor 30 Jahren – als unbestritten in der Lexikografie und Wörterbuchforschung angesehen werden, dass Wörterbücher als Gebrauchsgegenstände angesehen werden und dass deshalb die Benutzervoraussetzung (Wiegand 1998, 259; Wiegand et al. 2010, 680) die zentrale Größe bei der Planung lexikografischer Prozesse sein sollte (Bogaards 2003, 26, 33; Sharifi 2012, 626; Tarp 2008, 33–43; Wiegand 1998, 259–260).

Most experts now agree that dictionaries should be compiled with the users' needs foremost in mind. (Lew 2011a, 1)

So formulieren auch Bergenholtz und Tarp, dass eines der wichtigsten Ziele in der von ihnen begründete Funktionslehre ist, die Nutzer ins Zentrum zu rücken:

Consequently, all theoretical and practical considerations must be based upon a determination of these needs, i.e. what is needed to solve the set of specific problems that pop up for a specific group of users with specific characteristics in specific user situations. (Bergenholtz/Tarp 2003, 172)

Die Benutzungsforschung dient dabei nicht nur dazu, mehr über die Praxis der Wörterbuchbenutzung zu erfahren, sondern perspektivisch Wörterbücher auf Basis des daraus gewonnenen Wissens zu verbessern und benutzerfreundlicher zu gestalten.

Der Sinn der Benutzungsforschung, ihr forschungslogischer Status und ihre Legitimation ergeben sich daraus, daß Wörterbücher erarbeitet werden, um die Praxis ihrer Benutzung zu ermöglichen, und daß daher wissenschaftliche Kenntnisse zu dieser kulturellen Praxis eine der Voraussetzungen u.a. dafür sind, daß neue Wörterbücher [...] in dem Sinne benutzeradäquater sind, daß sie einen höheren Nutzungswert haben, wodurch sowohl die Voraussetzung für eine größere Benutzungseffizienz geschaffen wird als auch dafür, daß die Quote der erfolgreichen Benutzungshandlungen steigen kann. (Wiegand 1998, 259)

Doch auch wenn innerhalb der Wörterbuchbenutzungsforschung in den letzten Jahr(zeht)en sehr viel geforscht wurde, stellt Bogaards noch 2003 fest, dass “nevertheless, uses and users of dictionaries remain for the moment relatively unknown” (Bogaards 2003, 33). Dabei ist die Gruppe der nicht-muttersprachlichen Nutzer von Wörterbüchern noch der Bereich, der am besten erforscht ist:

Most progress in meta-lexicography has been made in relationship with L2 learners. Next to nothing is known when it comes to the use that is made of dictionaries by L1 users, or by the general public outside L2 courses. (ebd. 28)²

Daneben gibt es Studien zu kommerziellen Wörterbüchern von großen Verlagen (s. Hinweise bei Nesi 2012, 364), die allerdings meist als Geschäftsgeheimnis behandelt und demnach nicht veröffentlicht werden, sowie Studien zum Vergleich von gedruckten vs. elektronischen Wörterbüchern (cf. Dziemianko 2012) als auch weitere Studien aus dem universitären Bereich (zu einem Gesamtüberblick s. Welker 2010). Ein neuer Sammelband schließt einen Teil der Forschungslücke zu Onlinewörterbüchern, zu denen es bislang wenige Studien gab (Müller-Spitzer 2014a). Wenig gibt es allerdings aus dem Bereich der allgemeinen Nutzung von Wörterbüchern außerhalb von Schule, Lehre und Studium.

While it is not entirely true that lower-level educational stages have not been studied at all, there certainly is an imbalance in how well the dictionary-using population has been covered by past studies. In particular, dictionary use beyond the setting of educational institutions remains a severely underresearched black hole. (Lew 2011b, 5)

2 Methodisch-terminologische Grundlagen

In methodischer Hinsicht stützt sich die Benutzungsforschung vor allem auf die empirische Sozialforschung:

Will man die Methodologie eines neuen Forschungsgebietes entwickeln, dann muß dies nicht notwendig bedeuten, daß man sozusagen am ‚methodischen Nullpunkt‘ beginnt. Ein noch sehr junges Forschungsfeld wie die Benutzungsforschung ist vielmehr gut beraten, wenn es sich im näheren und weiteren Umfeld kundig macht und an die methodischen Erfahrungen anderer Forschungsfelder oder Disziplinen anknüpft. Hierzu gibt es derzeit mehrere Möglichkeiten. Nach meiner Auffassung [...] ist es für die Benutzungsforschung besonders förderlich, wenn sie auf Methoden der empirischen Sozialforschung und deren Erfahrung in der Methodenanwendung zurückgreift. [...] Die Übernahme von Methoden aus der empirischen Sozialforschung [...] geschieht nun allerdings nicht dadurch, daß innerhalb der Wörterbuchforschung eine erneute ausführliche Darstellung dieser Methoden erfolgt. Dies ist hier nicht beabsichtigt, und ich halte es auch für überflüssig; denn die Lehrbücher, Einführungen und speziellen Abhandlungen füllen bereits eine stattliche Bibliothek, die auch Wörterbuchforscher benutzen können. Die Adaption kann vielmehr in den künftigen Untersuchungen der Benutzungsforschung dadurch erfolgen, daß diese Methoden geprüft, erprobt, relativ zu Forschungsvorhaben gezielt übernommen und gegebenenfalls modifiziert und ergänzt werden. (Wiegand 1998, 568–569)

² Vgl. zu einem ähnlichen Statement auch Welker und zum Forschungsbedarf für Übersetzer Bowker (Bowker 2012, 380; Welker 2010, 10).

Entsprechend der Position, die in diesem Zitat deutlich wird, will auch der vorliegende Beitrag keine Übersicht über die unterschiedlichen Methoden bieten, die aus der empirischen Sozialforschung in die Wörterbuchbenutzungsforschung übernommen wurden. Vielmehr soll hier die Anwendung verschiedener Methoden aus dem Bereich der empirischen Sozialforschung anhand von Beispielen empirischer Wörterbuchbenutzungsstudien illustriert werden.

Diese Bandbreite an Methoden zu verdeutlichen, erscheint mir auch deshalb sinnvoll, weil die Diskussion um die methodische Qualität von Nutzungsstudien innerhalb der Metalexikografie teilweise sehr pauschalisierend geführt wird, wie die folgenden Beispiele zeigen:

But if one focuses exclusively on lexicography, it seems that almost no qualitative progress has been made. There are of course positive aspects, but it is not difficult to reach the conclusion that the majority of the previous user research is in fact 'a waste of time and money'. This holds especially true for the quantitative research projects. (Tarp 2009, 293)

[...] most of the studies of dictionary usage were carried out in the most unscientific way imaginable, as they were conducted without any knowledge and without the use of the methods of the social sciences. (Bergenholtz/Bergenholtz 2011, 190; für eine Zurückweisung dieser pauschalen Kritik s. Rundell 2012a, 3)

Allgemein kann man nur sagen, dass es in empirischen Erhebungen immer von der Forschungsfrage abhängt, welches Erhebungs- oder Forschungsdesigns man wählen sollte. Es gibt daher weniger gute und schlechte Designs generell, sondern eher angemessene und unangemessene bezogen auf eine Forschungsfrage. Genauso kann man auch nicht sagen, dass jede Befragung so gestaltet sein muss, dass die Befragten einen repräsentativen Querschnitt bilden.

A scientifically sound survey must be based on a section of the population surveyed that can be statistically viewed as representative of the entire population. (Bergenholtz/Bergenholtz 2011, 190)

In this way, they have broken the golden rule of sociology that informants should never select themselves or be selected by the researchers, but that the selection should always be made at random. Of course, there may be cases where the whole population of dictionary users are university students and where the random sample to be investigated is necessarily made up by students. But this is the exception to the rule in the 220 research projects, of which the majority has also broken another golden rule, i.e. that the number of informants should be large enough to provide statistically significant (and relevant) results. This lenient way of taking samples implies that they are not representative of the total population, and that the results cannot be generalised. It is nevertheless a fact that the researchers behind many of these projects do not hesitate to add percentages and decimals to everything that their miniature world of informants have done, looked for, wanted, etc. (Tarp 2009, 290)

Alle allgemeinen (Fragebogen-)Studien zur Wörterbuchbenutzung müssten nach dieser Position als Grundgesamtheit *alle* potentiellen Wörterbuchbenutzer haben, aus denen dann *immer* eine repräsentative Stichprobe gebildet werden müsste. Dies

wird aber nie möglich sein, da man die Grundgesamtheit aller Wörterbuchbenutzer nicht kennt. Bedeutet das nun, dass man keine solchen allgemeinen Studien durchführen sollte? Dies wäre m.E. eine ganz falsche Schlussfolgerung und würde auch nicht dem entsprechen, was in den empirischen Sozialwissenschaften oder auch in der Linguistik üblich ist.³ Vielmehr kann man sehr wohl eine Untersuchung mit einer nicht repräsentativen Stichprobe machen, man muss nur die Schlussfolgerungen entsprechend einschränken. Man kann eben von einer nicht repräsentativen Stichprobe nicht auf die Grundgesamtheit schließen, sehr wohl aber Aussagen über die jeweiligen Gruppen von Befragten treffen, was oft schon hinreichend ist. Befragt man beispielsweise 50 Übersetzer, 100 Sprachwissenschaftlerinnen und 70 Studierende, die keine Sprachwissenschaften studieren, so kann man aus diesen Stichproben gewisse Schlussfolgerungen für diese Gruppen ableiten. Man kann aber auf diese Weise natürlich keine Kenntnisse über *alle* Wörterbuchbenutzerinnen gewinnen.

Der folgende Beitrag zeigt verschiedene Studien auf, die unterschiedliche Erhebungsmethoden und -designs wie Onlinefragebögen, Logfileanalysen oder Eyetracking anwenden. Um diese Studien methodisch einordnen zu können, soll im Folgenden zunächst die relevante Terminologie für die einzelnen Schritte einer empirischen Erhebung kurz eingeführt und eingeordnet werden.

Für die Durchführung einer empirischen Untersuchung müssen in der Regel folgende Fragen geklärt werden (Diekmann 2011, 186–229; Koplenig 2014):

Wie kann ein Forschungsproblem formuliert und präzisiert werden?
Dieser Schritt bezieht sich auf die Ausarbeitung der *Forschungsfrage*.

Wie werden die entsprechenden Variablen gemessen?

Wenn die Forschungsfrage präzisiert und damit die theoretische Konzeption der Untersuchung geklärt ist, muss entschieden werden, wie man die einzubeziehenden Variablen messen will. Diesen Schritt nennt man *Operationalisierung*.

Welches Erhebungsdesign ist angemessen zur Erhebung der Daten?

Mit dem Erhebungsdesign wird der zeitliche Modus der Datenerhebung spezifiziert. Dabei werden drei Arten von Erhebungsdesigns unterschieden: (i) *Querschnittsdesign*, (ii) *Trenddesign* und (iii) *Paneldesign*. Ein (i) Querschnittsdesign bezeichnet eine Erhebung, bei der zu einem Zeitpunkt oder in einer kurzen Zeitspanne eine einmalige Erhe-

³ Man könnte nach dieser Prämisse beispielsweise keinerlei korpuslinguistische Untersuchungen machen, da auch diese keinen repräsentativen Ausschnitt aus der gesamten Sprache bilden können, da man diese Grundgesamtheit aus prinzipiellen Gründen nicht erfassen kann. Dies muss man natürlich im Auge behalten bezüglich der Aussagekraft solcher Untersuchungen; es stellt aber nicht grundsätzlich ihren Sinn in Frage.

bung mit beliebig vielen Probanden vorgenommen wird. Eine Querschnittserhebung erlaubt damit den Vergleich verschiedener Entitäten zu einem bestimmten Zeitpunkt. Von einem (ii) Trenddesign spricht man, wenn mehrere Querschnittserhebungen zum gleichen Thema zu mehreren Zeitpunkten durchgeführt wurden und diese zu einem Trend zusammengefasst werden. Genauer gesagt: Bei einem Trenddesign werden (a) die Werte der gleichen Variablen zu (b) mehreren Zeitpunkten mit (c) unterschiedlichen Stichproben, d.h. unterschiedlichen Probanden, erhoben. Im Unterschied zum Trenddesign werden im (iii) Paneldesign die a) Werte der gleichen Variablen zu (b) unterschiedlichen Zeitpunkten, aber mit (c) der gleichen Stichprobe, d.h. den gleichen Probanden, erhoben. Damit können im Paneldesign Entwicklungen auf individueller Ebene nachvollzogen werden. Trend- und Paneldesigns sind beides *Längsschnittdesigns*.

Welches Forschungsdesign ist hinsichtlich der Varianzkontrolle zur Beantwortung der Forschungsfrage am besten geeignet?

Mit der Entscheidung für ein Quer- oder Längsschnittdesign wird die zeitliche Dimension einer Datenerhebung spezifiziert. Bei der Planung einer empirischen Untersuchung ist noch ein weiterer Aspekt zu berücksichtigen, der sich auf die Bildung von Vergleichsgruppen und die Art der Aufteilung von Probanden auf diese Vergleichsgruppen bezieht. Dieser Aspekt wird auch als Varianzkontrolle bezeichnet. Dabei werden drei Arten von Designs unterschieden: (i) *experimentelle Designs*, (ii) *quasiexperimentelle Designs* und (iii) *Ex-post-facto-Designs*. In einem (i) experimentellen Versuchsdesign werden mindestens zwei Gruppen gebildet, wobei diese Gruppen nach einem Zufallsverfahren gebildet werden (Stichwort ‚Randomisierung‘) und die unabhängige Variable von der Forscherin manipuliert wird. Ein typisches Beispiel sind Medikamententests, in denen es eine Gruppe gibt, die ein Medikament bekommt, und dazu eine Kontrollgruppe, der ein Placebo verabreicht wird. Diese Gruppen bestehen demnach nur während des Tests. Für ein (ii) quasiexperimentelles Design gelten die gleichen Voraussetzungen wie für ein Experiment, nur mit dem Unterschied, dass die Bedingungen nicht zufällig verteilt werden, d.h. die Vergleichsgruppen werden zwar explizit und meist im Vorhinein im Rahmen der Untersuchungsplanung festgelegt, jedoch werden die Teilnehmerinnen den Vergleichsgruppen nicht zufällig zugewiesen. Ein Beispiel für ein solches Design können Mitarbeiterbefragungen sein, die vor und nach einer Umstrukturierung eines Unternehmens durchgeführt werden. Ein (iii) Ex-post-facto-Design ist ein Forschungsdesign ohne zufällige Zuordnung zu experimentellen Bedingungen und ohne Manipulation der unabhängigen Variablen, d.h. Gruppen von Probanden werden unterschieden auf Grund von Merkmalen, die auch schon vor der Untersuchung existierten und unabhängig von der Erhebung weiterbestehen. Dieses Design ist sehr üblich in Bezug auf Untersuchungen, die den Einfluss sozio-ökonomischer oder sozio-demografischer Faktoren auf Erziehung, Schulbildung oder beruflichen Erfolg untersuchen wollen.

Auf welche Weise sollen die Daten erhoben werden?

In der empirischen Sozialforschung werden vier Erhebungsmethoden unterschieden: (i) *Befragung* (persönlich, telefonisch, schriftlich), (ii) *Beobachtung*, (iii) *Inhaltsanalyse* und (iv) *Verhaltensspuren* oder *nichtreaktive Verfahren*. Die (i) Befragung ist die am häufigsten eingesetzte Erhebungsmethode in der Sozialforschung. Ein häufig angeführter Kritikpunkt ist dabei die Reaktivität der Methode, z.B. das Problem der sozialen Erwünschtheit. Damit ist gemeint, dass Probandinnen dazu neigen (können), Fragen so zu beantworten, wie es gesellschaftlich erwünscht ist. Das Problem der sozialen Erwünschtheit ist allerdings nicht für alle Lebensbereiche gleichermaßen relevant. Man kann zum Beispiel davon ausgehen, dass Fragen im Bereich gesellschaftlicher Toleranz stärker vom Effekt sozialer Erwünschtheit betroffen sind als Fragen in Bezug auf Wörterbuchbenutzung. (ii) Ganz allgemein sind sämtliche empirischen Methoden Beobachtungsverfahren, z.B. ist nur durch Beobachtung zu ermitteln, welche Stelle in einer Ratingskala angekreuzt wurde. Als Erhebungsmethode in der Sozialforschung ist mit Beobachtung jedoch spezifischer die direkte Beobachtung menschlicher Handlungen, sprachlicher Äußerungen, nonverbaler Reaktionen (z.B. Körpersprache) oder auch die Beobachtung sozialer Merkmale (Kleidung, Wohnungseinrichtung, Statussymbole) gemeint. Die (iii) Inhaltsanalyse befasst sich mit der systematischen Erhebung und Auswertung von Texten, Bildern und Filmen. Die Bezeichnung Inhaltsanalyse ist in einem gewissen Sinn zu eng, da auch formale Gesichtspunkte von Texten, z.B. die Länge von Sätzen, bei der Methode der Inhaltsanalyse eine Rolle spielen können. (iv) Nichtreaktive Verfahren sind solche, bei denen eine empirische Untersuchung ohne Wissen des Untersuchungsobjektes durchgeführt wird; eine Befragung ist dagegen ein Beispiel für eine reaktive Methode, da die Interviewsituation die Antwortreaktionen beeinflussen kann, da der oder die Befragte natürlich weiß, dass er/sie befragt wird. Diekmann gibt zur allgemeinen Unterscheidung reaktiver und nichtreaktiver Verfahren ein anschauliches Beispiel: Wenn man die Ernährungsgewohnheiten von Haushalten per Umfrage ermittelt, handelt es sich um eine reaktive Methode. Untersucht man dagegen mit der gleichen Zielsetzung die Haushaltsabfälle ohne das Wissen der ‚Probanden‘, so ist die Erhebungsmethode nichtreaktiv (Diekmann 2011, 195–196). Die Stärke von nichtreaktiven Verfahren liegt darin, unverfälschte Ergebnisse zu liefern und Daten über echtes Verhalten zu liefern. Gleichzeitig sind die Möglichkeiten zum Einsatz solcher Verfahren stark eingegrenzt, da die Forscher nur an wenigen Stellen Kontrolle über den Prozess haben.

Was ist für die Datenanalyse zu bedenken?

Sind Daten für eine empirische Forschung erhoben, müssen sie analysiert werden. Je sorgfältiger die vorangegangenen Schritte einer empirischen Erhebung durchgeführt wurden, desto besser wird die Datenanalyse funktionieren.

Manche Studie krankt daran, dass *irgendetwas* in einem sozialen Bereich untersucht werden soll, ohne dass das Forschungsziel auch nur annähernd klar umrissen wird. Auch mangelt es

häufig an der sorgfältigen, auf das Forschungsziel hin abgestimmten Planung und Auswahl des Forschungsdesigns, der Variablenmessung, der Stichprobe und des Erhebungsverfahrens. Das Resultat unüberlegter und mangelhaft geplanter empirischer ‚Forschung‘ sind nicht selten ein kaum noch genießbarer Datensalat und aufs äußerste frustrierte Forscher oder Forscherinnen. (ebd. 187; vgl. für die Lexikografie auch Lew 2011b, 8)

Was muss für die Berichterstattung beachtet werden?

Der letzte Teil einer empirischen Studie ist in der Regel die Berichterstattung. Grundsätzlich unterscheidet sich die Art der Berichterstattung empirischer Studien nicht von denen anderer Forschungsergebnisse. Allerdings hat sich ein bestimmtes Muster für die Darstellung empirischer Studien etabliert, welches in den meisten Veröffentlichungen angewandt wird: die sogenannte IMRAD-Struktur als Abkürzung für ‚introduction, method, results, and discussion‘ (Sollaci/Pereira 2004).

3 Das methodische Spektrum innerhalb der Benutzungsforschung verdeutlicht anhand einzelner Studien

Das Ziel des folgenden Abschnitts ist es, einige Beispiele von Studien kurz zu erläutern, die eine Bandbreite von Methoden anwenden. Dabei können die Untersuchungen nur in ihren Grundzügen dargestellt werden, meist mit nur einem kleinen Teil des eigentlich notwendigen Zahlenmaterials. Es werden aber immer Hinweise auf entsprechende Einzelpublikationen gegeben, in denen detailliertere Informationen zu den Analysen zu finden sind. Alle hier genannten Studien wurden am Institut für Deutsche Sprache in Mannheim durchgeführt.⁴ Die vier hier herausgegriffenen Forschungsfragen betreffen die Bildschirmgestaltung von Onlinewörterbüchern (3.1), die Einschätzung innovativer Merkmale in Onlinewörterbüchern (3.2), welche Stichwörter im deutschen Teil des Wiktionary auffällig oft nachgeschlagen werden (3.3) sowie die Frage, wie potenzielle Nutzerinnen zu Einzelbedeutungen in einem Onlinewörterbuch navigieren und wie sie Etikettierungen im Verhältnis zur Paraphrase rezipieren (3.4).

⁴ Die Studien sind zusammengefasst in Müller-Spitzer 2014a; s. auch <www.benutzungsforschung.de>. Sie alle sind das Ergebnis von Teamarbeit. Ich danke daher insbesondere Alexander Koplenig, der an allen hier berichteten Studien maßgeblich beteiligt war, und Sascha Wolfer, der die Logfile-Studie aus Abschnitt 3.3 (mit-)verantwortet hat.

3.1 Bildschirmgestaltung von Onlinewörterbüchern⁵

3.1.1 Einleitung

The challenge [...] is to try to assess which particular e-lexicographic solutions work best (and for whom, and under what circumstances), so that future electronic dictionaries can be made more effective than their paper predecessors, and more effective than the dictionaries available today. (Lew 2012: 344)

Dieses generelle Statement gilt für alle Aspekte von digitalen Wörterbüchern. Dementsprechend muss auch beim Design eines Onlinewörterbuchs beachtet werden, welches Design für welche Nutzergruppe am besten geeignet ist. Es ist bislang allerdings wenig zum grundlegenden Design von Wörterbüchern empirisch geforscht worden, außer z.B. bei Tono, der die Nützlichkeit von drei Interface-Gestaltungen testete und als Kontrollbedingung jeweils ein gedrucktes Wörterbuch benutzte (Tono 2000, vgl. auch Dziemianko 2012, 328). Einige Studien gibt es außerdem zu einer relativ speziellen Frage des Designs von Wörterbuchartikeln, und zwar zu ‚sign-posts‘ vs. ‚menus‘: Elementen, die die artikelinterne Suche nach Einzelbedeutungen in einem Artikel eines polysemen Stichworts erleichtern sollen (Lew/Tokarek 2010; Lew 2010; Nesi/Tan 2011); eine Studie setzt dabei auch Eyetracking als Form der Datenerhebung ein (Tono 2011).

Dabei stehen alle Wörterbuchprojekte, die keine etablierte Form der Datenpräsentation in elektronischen Wörterbüchern wählen wollen und die auch einfach neue Arten von Daten zu präsentieren haben, vor der schwierigen Frage, welche Form der Anordnung die beste ist. So berichten beispielsweise Trap-Jensen und Lorentzen über die Frage, wie Thesaurus-Daten im *Den Danske Ordbog*⁶ zu präsentieren seien:

This organization also reflects the editor's way of organizing the thematic group. There has, however, been heated discussion among the editors whether this is also the best way of presenting data. (Trap-Jensen/Lorentzen 2011, 177–178)

Diese Diskussionen hat es beispielsweise auch in *exlexiko* gegeben, einem wissenschaftlichen monolingualen Onlinewörterbuch zum Deutschen (vgl. Klosa et al. 2014), in dem die Wortartikel aus einem umfangreichen Angabespektrum bestehen und diese Angaben nicht einfach untereinander auf einem Bildschirm gezeigt, sondern sinnvoll aufgeteilt werden sollten. Auch in diesem Projekt wurde aber letztendlich nur projektintern diskutiert und eine Lösung gefunden, da keine passenden Wörterbuchbenutzungsstudien zu diesem Thema vorlagen. In *exlexiko* wurde beispielsweise eine Reiterdarstellung gewählt, in der die verschiedenen Angabebereiche auf verschiedene Bildschirmseiten verteilt werden. Der Vorteil dabei ist eine gute Übersichtlichkeit, ein Nachteil ist allerdings, dass der Überblick über den Gesamtartikel leicht

⁵ Diese Studie ist ausführlich dargestellt in Koplenig/Müller-Spitzer 2014.

⁶ <<http://ordnet.dk/ddo/>> [2014-02-21].

verloren geht. Damit wird die Atomisierung von sprachlichen Beziehungen, die an Wörterbüchern teilweise generell als kritisch angesehen wird, noch verstärkt:

Typically, the particularised presentation of lexical data in semasiological dictionaries, i.e. the individualised access to each lemma entry, does not bring the systematic nature of such phenomena to the fore, but rather obscures it by distributing the members of the set across the whole macrostructure. For some dictionary use situations, this is not a major issue, and some lexicographers counterbalance this effect by including systematic morphological or syntactic overview tables (inflection paradigms, inventories of closed class items, subcategorisation tables, etc.) into their dictionaries, for example as outer texts, in an appendix or in a dictionary grammar [...]. (Prinsloo et al. 2012, 297)

Diese Atomisierung ist also ein potentieller Nachteil der Reiteransicht. Auf der anderen Seite wird die Forderung, den Bildschirm nicht zu überfrachten, sondern ein dem elektronischen Medium angemessenes, gut lesbares Grunddesign zu bieten, eingelöst. Interessant ist daher die Frage, welche Aspekte potentielle Wörterbuchbenutzer bei verschiedenen möglichen Designs als positiv oder negativ herausstellen und wie sie verschiedene Ansätze bewerten. Eine Frage ist beispielsweise, ob die Übersichtlichkeit als zentraler Faktor erachtet wird, oder ob es als wichtiger angesehen wird, möglichst viele Informationen auf einen Blick zu sehen. Auch stellt sich die Frage, ob – wie es in dem Zitat von Lew herausgestellt wird – verschiedene Benutzergruppen unterschiedliche Bewertungen vornehmen. Eine Hypothese, die man vermuten könnte, wäre beispielsweise, dass Übersetzer, die in der Regel unter starkem Zeitdruck stehen, lieber einen schnellen Gesamtüberblick über den Artikel haben wollen und daher eine sehr verteilte Ansicht nicht präferieren, sondern eine, die den gesamten Wortartikel auf einer Bildschirmseite präsentiert:

Translators working in the twenty-first century are under enormous pressure to work quickly [...] for many translators, an ideal resource would be one that allowed them to do ‘one-stop-shopping’ [...] a translator in a working environment would benefit from having as much information as possible [...]. (Bowker 2012, 381)

Empirische Daten zu diesen Fragen des Grunddesigns zu erheben, würde in den Projekten solche oben beschriebenen Entscheidungen besser fundieren.

Daher haben wir versucht, empirische Daten zur Frage der Bildschirmgestaltung von Onlinewörterbüchern im Rahmen einer Onlinestudie zu erheben. Methodisch war diese Studie ein Querschnittsdesign, weil die Daten in einem Zeitschnitt erhoben wurden (11. August bis zum 16. September 2010). Die Studie war als Ex-post-facto-Design aufgebaut, d.h. in den Analysen nutzten wir demografische Daten wie beruflichen Hintergrund, Sprachversion, Alter etc. – Eigenschaften, die die Probandinnen unabhängig von der Befragung hatten. Die Form der Datenerhebung war eine Befragung in Form einer Onlinestudie. 390 Personen beendeten diesen Fragebogen.⁷

⁷ Die Rohdaten zu der Studie finden sich unter <www.benutzungsforschung.de>.

Das Ziel der Studie war es, Daten zu erheben, wie man einen Wortartikel mit einer detaillierten, ausführlichen Mikrostruktur, die auf einzelne Bildschirmteile oder -seiten aufgeteilt werden soll, am besten präsentieren kann. Um dies zu tun, wählten wir vier prototypische Möglichkeiten der Bildschirmgestaltung aus, die für wissenschaftliche Wörterbücher mit einem großen Angabespektrum bereits Anwendung finden. Wir wählten diese Art von Wörterbüchern aus, weil diese zum einen besonders von der Frage tangiert sind, wie sie ihre lexikografischen Daten präsentieren sollen (da es wenige Vorbilder gibt), und auf der anderen Seite für diese Art von Wörterbüchern noch keine Studiendaten vorliegen (ausgenommen von Bank (2010; 2012), die aber mehr auf allgemeine Fragen der Usability denn auf Design eingeht). Uns interessierte dabei nicht nur die Bewertung der einzelnen Ansichten, sondern auch vor allem auch die Gründe für diese Bewertung, da daraus dann allgemeinere Schlüsse bezüglich wichtiger Aspekte des Design gezogen werden können.

3.1.2 Methode

Die Probanden der Studie wurden gebeten, vier alternative Ansichten von Online-wörterbüchern zu bewerten und eine Ansicht auszuwählen, die sie präferieren, sowie Gründe für diese Präferenz zu nennen. Alle Ansichten enthielten den (fiktiven) Wortartikel „Sommer“ mit den (soweit wie möglich) identischen Angaben. Alle Beispiele außer der printorientierten Ansicht wurden mit Javascript implementiert.⁸ Daher konnten die Probanden interaktiv durch den Inhalt des Wortartikels im Onlinefragebogen navigieren.

Die erste Ansicht ist eine Adaption des bekannten Microsoft Windows Explorer, deshalb von uns auch ‚Explorer-Ansicht‘ genannt (vgl. Abb. 1). In diesem Grundlayout ist der Wortartikel wie ein Baum aufgebaut, dessen Teile selektiv auf- und zuklappbar sind, je nachdem, ob man auf das Plus- oder Minussymbol klickt. Zwei Beispiele, die dieses Grundlayout nutzen, sind das eben schon genannte *Den danske Ordbog* sowie das *Algemeen Nederlands Woordenboek*.⁹



Abb. 1: Explorer-Ansicht

⁸ Ich danke Peter Meyer für das Vorbereiten der relevanten Skripte.

⁹ <<http://anw.inl.nl/>> [2014-02-21].

In der zweiten zur Disposition gestellten Ansicht ist der Bildschirm in vier Teile, sogenannte *Panels* gegliedert (vgl. Abb. 2). Das *Digitale Wörterbuch der deutschen Sprache*¹⁰ setzt beispielsweise diese grundsätzliche Aufteilung des Bildschirms sein, wobei in diesem Wörterbuch nicht verschiedene Angabegruppen auf diese Panels verteilt sind, sondern unterschiedliche Ressourcen (einzelne Wörterbücher, Korpusbelege, Thesaurus etc.).



Abb. 2: Panel-Ansicht

Die dritte alternative Ansicht ist die sogenannte Tab-Ansicht (vgl. Abb. 3), in der man zwischen einzelnen Artikelgruppen in separaten Fenstern wechseln kann. Dieses Grundlayout wird sowohl in *elexiko*,¹¹ einem monolingualen deutschen Onlinewörterbuch, wie auch in *ELDIT*,¹² einem digitalen Lernerwörterbuch Deutsch–Italienisch eingesetzt.



Abb. 3: Tab-Ansicht

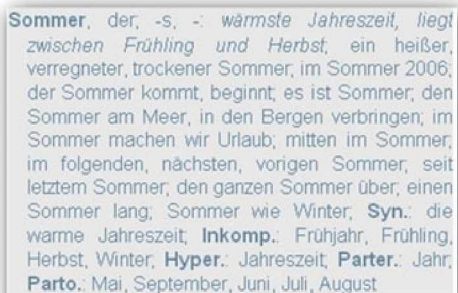
Die letzte Ansicht, die wir zur Bewertung gestellt haben, ist eine printorientierte Ansicht eines Wortartikels (vgl. Abb. 4), da es immer noch einige Onlinewörterbü-

¹⁰ <<http://www.dwds.de>> [2014-02-21].

¹¹ <<http://www.elexiko.de>> [2013-07-13].

¹² <<http://www.eurac.edu/eldit>> [2013-07-13].

cher gibt, die eng an ihre gedruckten Gegenstücke angelehnt sind, beispielsweise die Onlineversion des *Trésor de la Langue Française*.¹³



Sommer, der, -s, -: wärmste Jahreszeit, liegt zwischen Frühling und Herbst, ein heißer, verregneter, trockener Sommer; im Sommer 2006; der Sommer kommt, beginnt; es ist Sommer; den Sommer am Meer, in den Bergen verbringen; im Sommer machen wir Urlaub; mitten im Sommer; im folgenden, nächsten, vorigen Sommer; seit letztem Sommer; den ganzen Sommer über; einen Sommer lang; Sommer wie Winter; **Syn.:** die warme Jahreszeit; **Inkomp.:** Frühjahr, Frühling, Herbst, Winter; **Hyper.:** Jahreszeit; **Parter.:** Jahr; **Parto.:** Mai, September, Juni, Juli, August

Abb. 4: Print-Ansicht

Der Ablauf der Untersuchung in der Studie war folgendermaßen: Zunächst wurden den Teilnehmern die vier verschiedenen Bildschirmansichten präsentiert. Die Alternativen wurden in einer zufälligen Reihenfolge gezeigt, um Reihenfolgeeffekte (Diekmann 2011, 464–466) zu vermeiden. Nachdem die Probandinnen alle vier Beispiele gesehen hatten und ihre Benutzbarkeit austesten konnten, wurden sie gebeten, die Ansichten nach den Kriterien Qualität (1 = überhaupt nicht gut, 7 = sehr gut), Übersichtlichkeit (1 = überhaupt nicht übersichtlich, 7 = sehr übersichtlich) und Verständlichkeit (1 = überhaupt nicht verständlich, 7 = sehr verständlich) auf einer siebenstufigen Likert-Skala zu bewerten. Danach wurden die Teilnehmer gebeten, die Ansichten entsprechend ihrer Präferenz in eine Reihenfolge zu bringen (diejenige Ansicht, die am besten gefiel, als erstes usw.). Anschließend wurde den Probanden nochmals die Ansicht gezeigt, die sie als am besten eingestuft hatten und sie wurden gebeten, diese Einstufung in einer Freitextantwort zu begründen. Um potenzielle Gruppenunterschiede feststellen zu können, haben wir verschiedene Hintergrundvariablen wie akademischer und beruflicher Hintergrund oder die Sprachversion des Fragebogens in die Analyse einbezogen.

3.1.3 Ergebnisse und Diskussion

Die Tab-Ansicht war die Ansicht, die sowohl am besten bewertet wurde (vgl. Tab. 1) wie auch am häufigsten als beste Ansicht auf den ersten Platz gewählt wurde (vgl. Abb. 5). Auch wenn die Panel- und die Explorer-Ansicht etwas niedrigere Bewertungen auf einem aber insgesamt noch hohen Niveau bekamen, wurden sie seltener als beste Ansicht favorisiert. Die printorientierte Ansicht erhielt die schlechtesten Bewertungen und wurde auch am seltensten als beste Ansicht ausgewählt.

¹³ <<http://atilf.atilf.fr/>> [2013-07-13].

Alle Bewertungen der vier Alternativen wurden gemittelt, um einen zuverlässigen Maßstab der Bewertungen zu bilden, wobei höhere Werte eine bessere Bewertung anzeigen. Tabelle 1 fasst diese durchschnittlichen Bewertungen zusammen und zeigt, in wie viel Prozent der Fälle die jeweilige Ansicht auf den ersten Platz gewählt wurde.

Tab. 1: Mittelwerte und Standardabweichungen der Bewertungen sowie Erstrang-Prozente der vier alternativen Ansichten

Alternativen	Durchschnittliche Bewertung	Standardabweichung	Erstrang (Prozente)
Tab-Ansicht	5.43	1.39	42.82
Panel-Ansicht	5.15	1.46	32.82
Explorer-Ansicht	4.93	1.44	17.69
Print-Ansicht	3.36	1.55	6.67

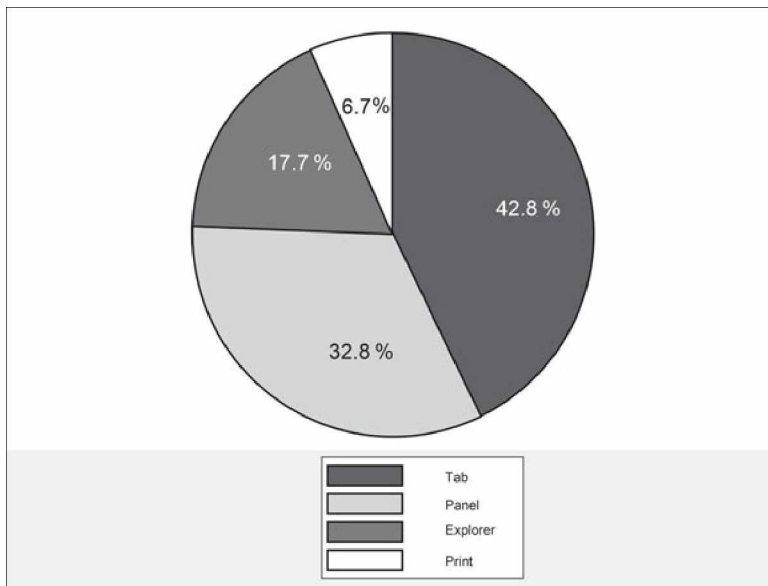


Abb. 5: Kuchendiagramm der präferierten Ansicht

Um herauszufinden, warum unsere Probanden eine bestimmte Ansicht favorisierten, haben wir manuell die Antworten auf die offene Frage („Hier sehen Sie noch einmal die Ansicht, die Ihnen am besten gefallen hat. Was gefällt Ihnen daran besonders gut?“) untersucht. Einige Teilnehmer begründeten ihre Auswahl relativ ausführlich, sodass wir hier interessante zusätzliche Daten über Vor- und Nachteile der verschiedenen Ansichten in der subjektiven Wahrnehmung unserer Teilnehmerinnen gewinnen konnten.

Die Antwortdaten haben wir nach dem in Tabelle 2 gezeigten Schema kodiert. Dabei haben wir bestimmte ‚typische‘ Formulierungen bestimmten Kategorien wie Übersichtlichkeit oder *Keine Notwendigkeit zu Klicken* zugeordnet, um anschließend in einem Bottom-up-Prozess analysieren zu können, welche wiederkehrenden Argumente für die Wahl einer bestimmten Ansicht angeführt wurden. Dies entspricht der Methode der Strukturierung (Diekmann 2011, 607–613; Mayring 2011), die wir in einem anderen Teil unserer Studien auch für die Analyse von Daten zu Benutzungskontexten angewendet haben. (Müller-Spitzer 2014b)

Tab. 2: Kodierungsschema zur Kodierung der Antworten auf die offene Frage

Kategorie	Auszüge aus Antworten
Übersichtlichkeit	alle notwendigen Informationen übersichtlich dargestellt. nicht „überladen“ uncluttered
Keine Notwendigkeit zu Klicken	alle Informationen sind übersichtlich und gut von einander abgetrennt aufgelistet ohne dass man weitere Mausklicks tätigen muss kein lästiges Herumklicken no clicking involved no need to click on anything
Keine Notwendigkeit zu Scrollen	ohne viel Scrollerei alle Info uebersichtlich auf dem Schirm doesn't force the user to scroll no need to scroll
Kein Overload an Information	aufgeräumt nicht zu viel Information auf einmal simple not too much information at once
Navigation	man sieht die verschiedenen Kategorien deutlich und kann gut zwischen ihnen wechseln easy to navigate easy to use
Look & Feel	optisch freundliche Anordnung modern Auge wird nicht durch viel zusammengequetschten Text überfordert keine farben, kein schnickschnack stylish visually appealing
Effizienz in der Benutzung	auf einen Blick klare, auf den ersten Blick erfassbare Gliederung functional intuitive
Adaptabilität/Selektivität	erlaubt mir, interessantes anzusehen und (fuer mich in dem moment) uninteressantes zu ignorieren it is possible to select only the information required adaptability of dictionary contents; I can choose

Kategorie	Auszüge aus Antworten
Wesentliche Information	die für mich relevante Information information unnecessary for me is not shown without sacrificing information to brevity hierarchical
Vertrautheit	Wahl zwischen den Informationsbereichen ist mir vertraut like the one I am used to similar to other applications consistent with web browser formatting
Schnelligkeit	Informationen, die für mich wichtig sind, kann ich zu jeder Zeit schnell abrufen schnell zu bedienen quick, open view presents all the data quickly does not take up traffic if used on a mobile phone
Anderes	
Weiß nicht/keine Antwort	

Tab. 3: Gründe für Präferenz (Prozente) einer Ansicht

Kategorie	Präferierte Ansicht					X ² / p-Wert ^a
	TAB	PANEL	EXPLORER	PRINT	Total	
Übersichtlichkeit	63.64^a	55.12	56.06	28.00	57.18	11.76 / 0.10
Keine Notwendigkeit zu Klicken	5.45	76.38	16.67	16.00	31.59	179.64 / 0.00
Navigation	36.12	29.13	25.76	16.00	40.81	5.92 / 1.00
Adaptabilität/Selectivität	32.12	3.94	56.06	8.00	25.33	71.68 / 0.00
Kein Overload an Information	24.24	8.66	16.70	28.00	18.54	13.30 / 0.05
Wesentliche Information	12.73	12.60	22.73	4.00	13.84	6.74 / 0.97
Effizienz in der Benutzung	10.30	11.02	9.09	12.00	10.44	0.24 / 1.00
Look & Feel	12.73	4.72	7.58	16.00	9.40	6.94 / 0.89
Vertrautheit	12.12	0.79	1.52	40.00	8.36	49.28 / 0.00
Schnelligkeit	5.45	6.30	7.58	12.00	6.53	1.67 / 1.00
Anderes	3.03	3.15	0.00	0.00	2.35	2.88 / 1.00
Keine Notwendigkeit zu Scrollen	1.82	0.79	0.00	0.00	1.04	2.00 / 1.00
Total	220.00	212.60	222.73	180.00	215.40	

^a Die drei meistgenannten Kategorien für jede Ansicht sind in fett dargestellt. Die p-Werte sind Bonferroni angepasst.

In Tabelle 3 wird die Häufigkeitsverteilung der einzelnen Kategorien für jede Ansicht gezeigt. Die drei meistgenannten Argumente für die jeweilige Ansicht sind dabei fett hervorgehoben. Übersichtlichkeit ist dabei das am häufigsten genannte Kriterium für die Wahl einer Ansicht (57.18%) besonders bei den Teilnehmern, die die Tab-Ansicht bevorzugten (63.64%). Verglichen zur Panel-Ansicht (3.94%) und zur printorientierten Ansicht (8.00%) wird sowohl die Tab- (32.12%) wie auch die Explorer-Ansicht (56.06%) deshalb präferiert, weil sie sehr adaptiv erscheint. Dieser Unterschied ist hochsignifikant.¹⁴ Eine Benutzeroberfläche, in der man gut navigieren kann, scheint auch ein wichtiger Grund für die Auswahl einer Ansicht zu sein, insbesondere bei denen, die die Tab- (36.12%), die Panel- (29.13%), oder die Explorer-Ansicht (25.76%) wählten. Im Vergleich zu den anderen drei Möglichkeiten der Bildschirmgestaltung wird die Panel-Ansicht (76.38%) vor allem deshalb präferiert, weil sie Nutzern erlaubt, alle Informationen auf einen Blick zu erhalten.¹⁵ Für die Auswahl der printorientierten Ansicht wird vor allem herausgestellt, dass sie vertraut erscheint (40.00%). Der Unterschied zu den drei Alternativen ist auch wiederum hochsignifikant.¹⁶ So konnten wir anhand dieser Antwortdaten Indikatoren dafür gewinnen, welche Gründe zu der Einstufung der Ansichten führten.

Zusammenfassend zeigen unsere Analysen, dass die meisten Probandinnen die Tab-Ansicht präferieren. Gruppenspezifische Unterschiede (hier nicht weiter berichtet) scheinen dabei keine wesentliche Rolle zu spielen. Weder die gewählte Sprache des Fragebogens (deutsch/englisch), noch beruflicher oder sprachwissenschaftlicher Hintergrund, noch das Alter der Teilnehmer beeinflussen die Verteilung signifikant: in fast jeder Subgruppe liegt die Tab-Ansicht auf dem ersten Platz. Die Analyse der Antwortdaten auf die offene Frage zeigt, dass die Probandinnen diese Form der Bildschirmgestaltung präferieren, weil sie übersichtlich, einfach zum Navigieren und adaptiv ist. Eine Ausnahme bilden dabei nur die Übersetzer.¹⁷ Diese bevorzugen die Panel-Ansicht. Damit kann die anfangs aufgestellte Hypothese, dass Übersetzerinnen eine Ansicht bevorzugen könnten, die alle Daten auf einen Blick liefert, im Rahmen unserer Untersuchung als bestätigt gelten. Die Unterschiede sind allerdings auch hier nicht hoch, sodass man dieses Ergebnis nicht überbewerten sollte.

Man kann aus den Daten nicht schließen, dass die Tab-Darstellung in tatsächlichen Benutzungssituationen die präferierte ist, weil z.B. der Nachteil der fehlenden Übersichtlichkeit in der Fragebogensituation nicht in gleicher Weise zum Tragen kommt wie in einer tatsächlichen Benutzungssituation (d.h. ein „Benutzer-in-actu“, wie Wiegand es ausdrücken würde, könnte andere Kriterien anlegen, Wiegand 1998, 501). Es ist eher eine Einschätzung zur Gefälligkeit des Grunddesigns, die hier

¹⁴ $\chi^2(3) = 71.68, p < .005$

¹⁵ $\chi^2(3) = 179.64, p < .005$

¹⁶ $\chi^2(3) = 49.28, p < .005$

¹⁷ Zu genaueren Ergebnissen s. Koplenig/Müller-Spitzer 2014, 196.

erhoben wurde. Allerdings zeigen die Antworten auf die offene Frage sehr deutlich, dass den Probanden diese Vor- und Nachteile grundsätzlich auch in der Studie deutlich wurden. Beispielsweise ist es ein oftmals wiederkehrendes Argument für die Wahl der Panel-Ansicht, dass man dort alles auf einmal sehen kann, wie z.B. in der folgenden Antwort:

- The information is well-ordered. All sections of the entry can be viewed either simultaneously or separately (which is what the view with tabs cannot do).

Genauso schreibt jemand, der sich für die Tab-Ansicht entschieden hat:

- most intuitive online – can have as clear a page as you want. Unlikely to be comparing the different tabs at the same time, takes few clicks to navigate around.

Diese Kritikpunkte an der Tab-Ansicht werden von einem Teilnehmer sogar durch Verbesserungsvorschläge ergänzt:

- Although it hides some information, and requires excess clicking, the uncluttered, tabular interface helps focus your attention on the details you are looking for. If this were paired with a customizable search that brought you to the tab corresponding most to what you're searching for (e.g. 'dog' sense relations → sense relations tab for entry 'dog'), this would be fantastic.

Auch werden Vergleiche zwischen den verschiedenen Ansichten gezogen, die zeigen, dass viele grundlegende Eigenschaften auch in der Fragebogensituation zu erkennen waren:

- That all the relevant information is on one page, immediately visible without further clicking. The two-dimensional arrangement without any visible boxes is somewhat irritating and the categorization of the examples is missing, which is a pity. The tree structure was OK, but having to explicitly open not only the first, but also the second level was a bit much. The article in print dictionary style would have been fine, too, if line breaks and paragraphs were inserted, the abbreviations spelled out, and all information from, for example, the tabbed version available. In this tabbed version you always have to click back and forth, and are never able to see the data side by side.

In Abschnitt 2 wurde der Effekt der sozialen Erwünschtheit erwähnt. Man könnte auch in diesem Zusammenhang argumentieren, dass der hohen Bewertung der Tab-Ansicht eine Verzerrung aufgrund sozialer Erwünschtheit zugrunde liegt. Da das Projekt, in dem wir die Studien durchgeführt haben, eng mit Elexiko verbunden war und dieses Wörterbuch die Tab-Ansicht einsetzt, könnten die Teilnehmerinnen gedacht haben, dass dies eine von uns erwünschte Antwort wäre. Dieser Frage sind wir nachgegangen und konnten diese Einschränkung entkräften, da es keinen signifikanten Zusammenhang zwischen Sprachversion des Fragebogens und der Präferenzdistribution gab. Elexiko ist ein monolinguales deutsches Onlinewörterbuch, sodass es als sehr unwahrscheinlich gelten kann, dass es den englischsprachigen Probanden so sehr

bekannt ist, dass ihr Antwortverhalten vom Effekt sozialer Erwünschtheit verzerrt ist. Außerdem wissen wir aus einer anderen Fragebogenstudie, die nur auf Deutsch erhoben wurde, dass elexiko nur rund 21% der deutschsprachigen Teilnehmerinnen bekannt war (Klosa et al. 2014, 288–289). So kann man in manchen Fällen auch mehr oder weniger sicher prüfen, ob das Frageverhalten verzerrt ist.

Die Analyse der offenen Frage zeigt sehr deutlich die Gründe für die Präferenzen. Überraschend ist dabei, dass die Adaptivität ein sehr oft genanntes Kriterium ist. Dies ist deshalb überraschend, weil dieses Kriterium in einer anderen von uns durchgeführten Studie bei den Merkmalen guter Onlinewörterbücher als sehr unwichtig bewertet wurde (Müller-Spitzer/Koplenig 2014, 148–149). Weniger überraschend ist, dass das Kriterium der Übersichtlichkeit häufig als wichtig herausgestellt wird. Dies deckt sich mit den allgemeinen Erhebungen zu den Merkmalen guter Onlinewörterbücher aus der eben genannten anderen Studie. Die Bandbreite dessen, was als übersichtlich angesehen wird, ist allerdings sehr groß. M.E. kann das, was Probanden in der Hinsicht schreiben, nicht als Argument für monofunktionale Wörterbücher angesehen werden, wie es beispielsweise Bergenholtz/Bergenholtz vorschlagen (2011; vgl. auch Bergenholtz/Bothma 2011, 54–57; Bergenholtz 2011, 53). Denn ein Aspekt von Übersichtlichkeit, den die Teilnehmerinnen hervorhoben, ist auch, dass die Darstellung nicht überfrachtet, aber doch auch eine Vielzahl von Daten zu sehen sein soll.

- I like to see everything at once, but I like it separated into categories.
- It has everything clearly presented. I dont have to keep clicking on more options to find out more info. It's all right there.

Auch zeigte sich an anderer Stelle in unseren Studien, dass die Probanden im Zusammenhang mit einer benutzeradaptiven Anpassung jede Form von Profilwahl als besonders problematisch ansehen. Als ein Beispiel:

- please do not make the user have to select a whole bunch of things before getting to the dictionary entry. This would be a fatal choice and make the dictionary annoying and difficult to use. People would choose to use a dictionary, which is qualitativ worse but easier to use over the one where you have to fill in a whole bunch of baloney before you use it! People want answers fast! And then they want to play around with them. We are not all scientists who search for information systematically.

Dies ist auch ein Argument gegen die theoretisch überzeugende Idee eines *Entscheidungsbaums*, der in mehreren Schritten zu den lexikografischen Angaben führen soll, die man in einer bestimmten Situation benötigt (Prinsloo et al. 2012, 308–309). All diese Ideen bescheren dem Benutzer zwar am Ende einen passenden Wörterbucharartikel, aber der Weg dahin ist sehr lang und man muss sich fragen, welche Benutzer gewillt sind, diese Hürde zu überspringen, wenn schon ein notwendiges Login manche Benutzerinnen davon abhält, ein Wörterbuch zu benutzen (vgl. Bank 2012, 356–357).

Die hier vorgestellten empirischen Daten über Fragen des Designs geben den Projekten, die die oben genannten oder ähnliche Formen von Ansichten für ihre lexikografischen Daten verwenden, insgesamt wertvolle Hinweise dazu, wie eine Auswahl potentieller Wörterbuchbenutzer diese einschätzen und bewerten. Darüber hinaus zeigen die Antworten auf die offene Frage auch in etwas abstrakterer Form und losgelöst von konkreten Designmodellen, welche Kriterien diese potentiellen Benutzerinnen als besonders wichtig hervorheben. Die Übersichtlichkeit scheint sich durch viele Antworten als Leitaspekt hervorzutun, wie auch eine mögliche adaptive Anpassbarkeit, wenn sie nicht eine zu komplexe Bedienungsführung bedingt. Ein wichtiger, immer wiederkehrender Punkt ist die intuitive Bedienbarkeit eines Onlinewörterbuchs. Dies trifft auch für ganz andere Bereiche zu, wie beispielsweise ein Interview mit Rüdiger Grube, dem Vorstandsvorsitzenden der Deutschen Bahn AG zeigt, der – befragt zu Prognosen für das Mobilitätsverhalten der Zukunft – antwortet:

Was sich am Mobilitätsverhalten der Deutschen mit Sicherheit nicht ändern wird, ist, dass alles einfach sein muss und möglichst bequem. Mobilitätsangebote, für die man ein Gelehrter sein muss, um sie zu verstehen, haben keine Zukunft.¹⁸

So hat es ein Teilnehmer sicher auch für das Design von Onlinewörterbüchern auf den Punkt gebracht, wenn er herausstellt:

- Wenn ich eine Einführung brauche, hat das Layout versagt.

Die eben beschriebene Studie macht m.E. auch deutlich, dass Bergenholtz/Johnsen in dem folgenden Zitat keine Schwäche von Befragungen als Erhebungsmethode an sich thematisieren, sondern nur ein Beispiel für eine schlechte Befragung nennen.

In such questionnaire surveys, the same methods are employed as in other forms of market analysis: a number of standard questions or behavior, e.g. of a selected sample concerning a certain product or behavior (...) However, the answers from the informants do not necessarily reflect a real genuine user situation. It cannot be ruled out that the problems, behavior, etc. described by the informants differ from their real problems. The questions asked have to do with future activities, as in “Under which headword would you look for the following collocations?”, or with past activities “Which types of information do you look for most often”. There is no guarantee that the answers correspond to why and how the informants really have used or will use dictionaries. Such surveys are quite problematic because they presuppose that the informants remember exactly how they have used dictionaries in the past and that they are able to predict how they will do it in the future. (Bergenholtz/Johnsen 2005, 119–120)

Forschungsfragen wie die eben beschriebene kann man dagegen sehr gut mit in einer Befragung untersuchen.

¹⁸ Interview mit Peter Ramsauer und Rüdiger Grube, DB mobil 9/2012: 45.

3.2 Die Einschätzung innovativer Features von Onlinewörterbüchern

3.2.1 Einleitung

Innovative Features wie die Einbindung multimedialer Daten, die benutzeradaptive Anpassbarkeit dessen, was in einer bestimmten Benutzungssituation an lexikografischen Daten relevant ist oder flexiblere Zugriffsstrukturen auf die Wörterbuchinhalte gelten als besondere Pluspunkte digitaler Wörterbücher. Dies wurde von (Meta-)Lexikografinnen auch schon früh erkannt (de Schryver 2003). Allerdings werden solche Features – wenn sie tatsächlich geboten werden – bislang nicht in dem Maße genutzt, wie die Lexikografen es sich wünschen würden:

Whether they adhere to one school of thought or another, most lexicographers welcome the possibility of showing exactly the relevant information categories in a particular lookup situation, no less and no more, tailored to the specific needs and skills of the user. For the lexicographer, this is a strong argument in favour of the e-dictionary over the printed dictionary: the electronic medium has solved some of the problems related to traditional dictionaries. For the same lexicographers, it may be disappointing that the users do not seem to take advantage of all these wonderful possibilities. (Trap-Jensen 2010, 1142)

Dies deckt sich mit der eben schon zitierten empirischen Studie, in der wir nach Kriterien guter Onlinewörterbücher gefragt haben (Müller-Spitzer/Koplenig 2014). Diese Onlinefragebogenstudie, die 684 Menschen vollständig ausfüllten, brachte das Ergebnis, dass im Gegensatz zu klassischen Eigenschaften guter Referenzwerke (inhaltliche Verlässlichkeit, Übersichtlichkeit) die medienspezifischen Möglichkeiten für digitale Wörterbücher (multimediale Features, benutzeradaptive Anpassung), als unwichtig eingestuft wurden. Dies ist auf der einen Seite nicht überraschend, da z.B. ein multimedial sehr gut aufbereitetes, aber inhaltlich nicht verlässliches Referenzwerk kaum nutzbringend ist. Es zeigte sich in der Studie aber darüber hinaus, dass diese Einstufungen nicht nur in Konkurrenz, sondern auch unabhängig voneinander vorgenommen wurden: Die Teilnehmerinnen stufen die Kriterien Multimedialität und Adaptivität grundsätzlich als unwichtig ein, nicht nur im Verhältnis zu den klassischen Eigenschaften guter Referenzwerke. Der Deutungsansatz, dass die unterschiedliche Bewertung in der Konkurrenz der Eigenschaften zu suchen ist, reicht daher nicht aus. Ein anderer Interpretationsansatz ist, dass unsere Teilnehmer vielleicht noch nicht genug sinnvolle Beispiele solcher innovativer Features kennen.

[...] people are not born with the skills to extract the wealth of data stored in dictionaries and other reference works efficiently and transform it into knowledge. It takes time to get accustomed to new ways of finding information, it may even require formal training. (Trap-Jensen 2010, 1142; vgl. auch Heid/Zimmermann 2012, 669; Tarp 2011, 59; Verlinde/Peeters 2012, 151)

Insofern ist es auch möglich, dass unsere Probandinnen aus diesem Grund den potentiellen Nutzen innovativer Features nicht vollumfänglich einschätzen können. Diese These haben wir in Form eines Experiments empirisch überprüft. Das Experiment (als eine Art von Forschungsdesign) war dabei in eine Querschnittserhebung in Form einer Onlinefragebogenstudie¹⁹ integriert.

Die Forschungsfrage war, ob Probanden die Nützlichkeit von multimedialen Features oder Möglichkeiten benutzeradaptiver Anpassung in Onlinewörterbüchern als besser einschätzen, wenn sie vorher über Möglichkeiten dieser Features informiert wurden. Unsere Hypothese war dabei, dass Probandinnen, die anhand von Beispielen über die Möglichkeiten solcher Features vorher informiert werden, deren Nützlichkeit als höher bewerten, weil sie aus ihrer alltäglichen Wörterbuchpraxis wahrscheinlich zu wenige Beispiele kennen, um ohne diese Demonstration wirklich abschätzen zu können, wie hilfreich solche Elemente sein können.

3.2.2 Methode

Um diese Hypothese zu prüfen, haben wir ein Experiment in eine Onlinestudie (N=390) integriert.²⁰ In einem ersten Set (S1) zeigten wir die Möglichkeiten von multimedialen und benutzeradaptiven Features. Jedes Feature wurde textuell erläutert und anhand einer Abbildung illustriert (s. zwei Beispiele in Abbildungen 6 und 7). Die Teilnehmer wurden daraufhin gebeten, jedes einzelne Feature in Hinsicht auf drei Kriterien (wichtig/nützlich/hilfreich) auf einer 7-stufigen Likert-Skala zu bewerten. In einem zweiten Set (S2) wurden die Teilnehmerinnen gefragt, inwieweit sie den folgenden Aussagen zustimmen:

- (A) Die multimediale und benutzeradaptive Gestaltung eines Onlinewörterbuchs ist eine Arbeitserleichterung.
- (B) Die multimediale und benutzeradaptive Gestaltung eines Onlinewörterbuchs ist eine technische Spielerei.

Um einen potenziellen Lerneffekt untersuchen zu können, randomisierten wir die Reihenfolge dieser Sets. Den Teilnehmer der Versuchsgruppe, d.h. denen der Lerneffekt-Bedingung, wurden zunächst die Beispiele aus S1 gezeigt. Danach wurden sie gebeten, ihre Einschätzung der Nützlichkeit dieser Features in S2 zu geben. Die Teilnehmerinnen der Kontrollgruppe mussten dagegen S2 vor S1 beantworten, d.h.

¹⁹ Dies ist ein Beispiel dafür, wie wichtig es ist, zwischen den verschiedenen Eigenschaften einer empirischen Erhebung zu unterscheiden. Mit „Experiment“ ist nämlich nicht notwendigerweise ein Labortest gemeint, sondern dieses Forschungsdesign kann auch wie hier in Befragungen integriert werden.

²⁰ Diese Fragebogenstudie war die gleiche wie in Abschnitt 3.1, d. h. lief von 11.8. – 16.9.2010 (N = 390).

die Probandinnen der Lerneffekt-Bedingung hatten eine andere Basis – mehr Vorwissen – für ihre Entscheidung als die Teilnehmer der Nicht-Lerneffekt-Bedingung. Sollte unsere Hypothese sich bewahrheiten, müssten im Ergebnis die Teilnehmerinnen der Lerneffektbedingung adaptive und multimediale Features als nützlicher einschätzen als die Teilnehmer der Nicht-Lerneffekt-Bedingung.

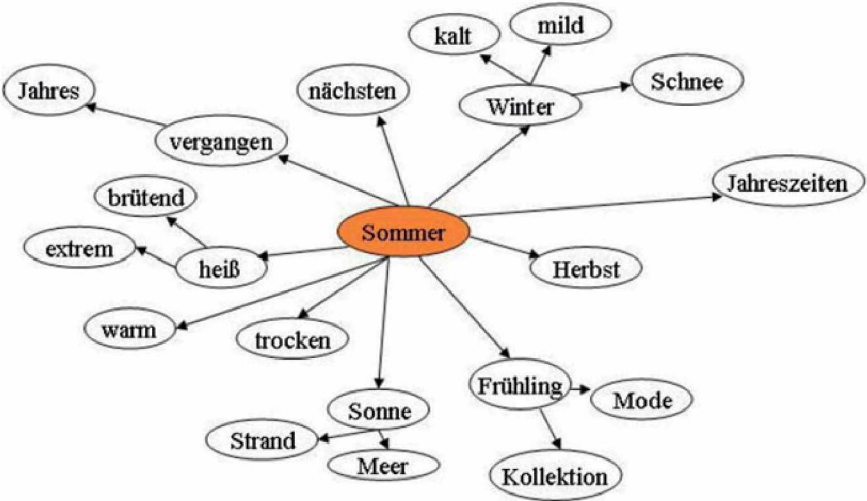


Abb. 6: Beispiel eines multimedialen Features im Fragebogen



Abb. 7: Beispiel eines adaptiven Features im Fragebogen

3.2.3 Ergebnisse und Diskussion

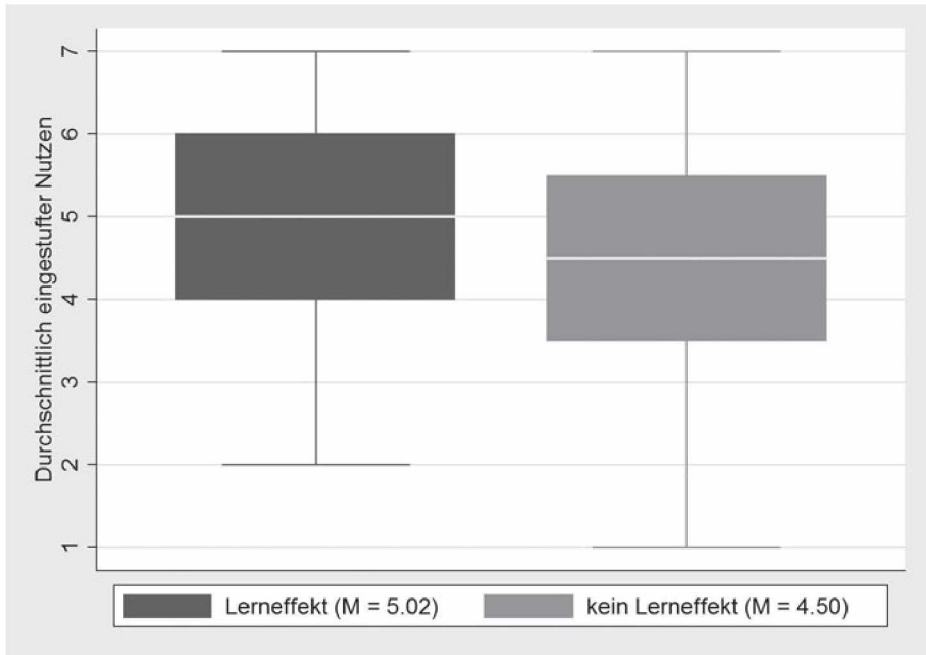


Abb. 8: Boxplots: Bewertung von multimedialen und adaptiven Features abhängig von der Lerneffekt vs. Nichtlerneffekt-Bedingung

Im Ergebnis zeigte sich, dass die Teilnehmer der Versuchsgruppe die Nützlichkeit der Features als signifikant höher bewerteten als die der Kontrollgruppe (vgl. Abb. 8). Die Grafik ist dabei folgendermaßen zu lesen: Die Probandinnen wurden gebeten, die Nützlichkeit der Features in einer siebenstufigen Likertskala (1 = stimme überhaupt nicht zu, 7 = stimme voll und ganz zu) zu bewerten. Diese Zahlen finden sich auf der y-Achse. In den Boxplots sieht man die Streuung der Daten. Die ausgefüllten Kästen entsprechen dem Bereich, in dem die mittleren 50% der Daten liegen. Der Lerneffekt, der sich hier zeigt, ist zwar moderat, aber hochsignifikant, d.h., die empirischen Daten stehen in signifikantem Gegensatz zur Nullhypothese, welche annimmt, dass zwischen den experimentellen Gruppen kein Zusammenhang besteht.

Unsere Hypothese konnte in diesem Experiment also bestätigt werden: Probandinnen, die zunächst Beispiele innovativer Features gezeigt bekamen, stuften diese als nützlicher ein als die Probanden, die diese Beispiele nicht gezeigt bekamen. Aus unseren Daten kann man daher schließen, dass es lohnend ist, innovative Features in Onlinewörterbücher zu integrieren, dass man aber gleichzeitig das Potenzial dieser Features den Nutzerinnen auch in irgendeiner (aktiven) Weise nahe bringen muss. Die Frage ist allerdings, wie dies in der Praxis aussehen kann, da die Lexikografen

in der Regel keinen direkten Kontakt zu ihren Nutzerinnen haben. Eine Möglichkeit könnte sein, geschlossene Kontexte in Bildungseinrichtungen wie den schulischen und universitären Unterricht stärker zu nutzen, in dem man Kontakt zu Nutzerinnen aufbauen kann mit der Chance, diese Nutzer zu schulen. Damit wird man sicher nicht solche Benutzer erreichen, die schnell die Rechtschreibung eines Stichworts prüfen wollen, aber vielleicht solche, die an weitergehenden Formen von Wörterbuchbenutzung interessiert sind.

Abschließend noch einmal zu einer methodischen Diskussion: Experimente gelten in den empirischen Wissenschaften als das Forschungsdesign, in dem es am besten gelingt, den Einfluss sämtlicher Drittvariablen (auch der unbekannten) zu neutralisieren (Diekmann 2011, 339). Allerdings haben Experimente auch Nachteile, wie z.B. das „Problem der externen Validität der geprüften Zusammenhänge“ (ebd. 349). Deshalb ist es wiederum eine zu pauschale Einschätzung, Experimente als den Königsweg der empirischen Wissenschaften einzuordnen:

[...] most psychologists consider experiments to be the ‘royal road’ of empirical research. The advantage of experiments is that it is possible to measure how the introduction of certain types of data or groups of such data in dictionaries and other lexicographical tools influence the result of the consultation by various types of users. A disadvantage is that they are very time-consuming and require a large number of informants to be representative, and that it is necessary to have knowledge of statistic methods or collaborate with statisticians with a view to determining whether the research results are statistically significant. (Tarp 2009, 288)

Wichtig scheint mir aber vor allen Dingen festzuhalten, dass dieses Zitat zeigt, dass in der Wörterbuchbenutzungsforschung oft nicht hinreichend genau zwischen Erhebungs- und Forschungsdesigns unterschieden wird (Koplenig 2014, 61–62). Ein experimentelles Design ist – wie das eben erläuterte Beispiel gezeigt hat – eine Art des Forschungsdesigns, welches auch beispielsweise in eine Befragung integriert werden kann. Man kann daher nicht sagen, dass ein Experiment an sich besonders zeitaufwendig ist oder pauschal eine hohe Anzahl an Probandinnen verlangt. Im Übrigen gilt für *alle* empirischen Untersuchungen, dass man interdisziplinär mit Fachleuten zusammen arbeiten sollte, die die notwendigen statistischen Kenntnisse haben, um alle Schritte einer empirischen Erhebung fachlich angemessen durchführen bzw. begleiten zu können.

3.3 Auffällig oft nachgeschlagene Stichwörter im deutschen Wiktionary

3.3.1 Einleitung

Die Analyse von Logfiles, d.h. von Protokolldateien, die den Zugriff auf digitale Wörterbücher aufzeichnen, wird als vielversprechende neue Form der Datenerhebung in der Wörterbuchbenutzungsforschung angesehen (u.a. Bergenholtz/Johnsen 2005; de Schryver et al. 2006; de Schryver/Joffe 2004). Die Analyse von Logfiles gehört von der

Erhebungsmethode in die Gruppe der beobachtenden Verfahren und hat damit sowohl das Potenzial (z.B. die Aufzeichnung echten Verhaltens im natürlichen Kontext) als auch die Einschränkungen (z.B. geringe Kontrolle des Prozesses durch die Forscher) dieser Erhebungsmethode (vgl. auch Diekmann 2011, 548ff.). Aus diesem Grund argumentiert Lew im folgenden Zitat, dass es nicht sinnvoll sei, aus den Logfiles unmittelbare Folgerungen für die weitere Bearbeitung eines Wörterbuchs zu ziehen:

If used in the latter sense, log files have a number of limitations. The context of dictionary use is completely unknown (with the possible exception of the user having been redirected from a well-known site, as indicated by *http referrer logging*). Furthermore, we typically know precious little about the user. We cannot be sure that the user has selected an even remotely appropriate tool for the job. For example, if log files show that someone has typed in *Powerpuff Girls* into our online dictionary, what do we do with this information? For all we know, this could be an 8-year old trying to print a colouring page of her favourite cartoon characters. So where do we go from here? Should we let ourselves be tempted to modify the dictionary to dutifully and indiscriminately serve all types of oddball queries? Whatever happened to the mantra of *genuine purpose* here? And how is this better than letting user questionnaires shape the dictionary: an approach criticized (among others) on the grounds that users tend to be limited by their particular experience, imagination, and preconceptions of what a dictionary is supposed to be like? Leaving these questions in the air for now, I would like to urge the researchers employing online log files to reflect on the limitations of this admittedly convenient data source. (Lew 2011b, 7)

Im Folgenden soll nun ein Beispiel für eine Studie basierend auf der Analyse von Logfiles vorgestellt werden, die m.E. die Stärken nutzt, ohne zu viele Schlussfolgerungen aus der Analyse zu ziehen. Diese Studie ist in methodischer Hinsicht auch deshalb interessant, weil sie nicht nur eine relativ neue Form der Datenerhebung nutzt, sondern auch vom Erhebungsdesign im Gegensatz zu den anderen hier vorgestellten Studien als Trenddesign und damit als Längsschnittstudie einzuordnen ist.

Die Forschungsfrage, der wir in dieser Studie nachgegangen sind, ist, welche Wörter im deutschsprachigen *Wiktionary*²¹ auffällig oft nachgeschlagen werden (Wolfer/Koplenig/Müller-Spitzer erscheint). Diese Studie ist eine Anschlussstudie zu einer Analyse, ob häufige Wörter im Korpus allgemein auch häufiger nachgeschlagen werden (Koplenig et al. 2014). Diese letztgenannte Studie war eine Antwort auf die zunächst von de Schryver gestellte Frage „Do dictionary users really look up frequent words?“ (de Schryver et al. 2006), in der im Gegensatz zu der de-Schryver-Studie anhand der Logfiles des *Digitalen Wörterbuchs der deutschen Sprache*²² und des deutschen Wiktionary nachgewiesen werden konnte, dass es einen signifikanten Zusammenhang zwischen Korpushäufigkeit und Nachschlagehäufigkeit gibt. Im Anschluss haben wir uns gefragt, welche anderen Faktoren – außer der Frequenz im Korpus – die Nachschlagehäufigkeit beeinflussen.

²¹ <<http://de.wiktionary.org/>> [2014-02-21].

²² <www.dwds.de> [2014-02-21].

Um dies zu untersuchen, analysierten wir die Logfiles des deutschen Teils des Wiktionary von Januar bis Dezember 2013. Die Wikimedia Foundation²³ veröffentlicht stündlich Statistiken über alle Aufrufe von Seiten, die zur Wikimedia Gesellschaft gehören (wie *Wikipedia* oder *Wiktionary*).²⁴ Das deutschsprachige Wiktionary ist in den Logdateien beispielsweise mit „de.d“ gekennzeichnet. Jeder Logfileeintrag verzeichnet dabei den Titel der aufgerufenen Seite, den Name des Wikimedia-Projekts, den Anzahl der Zugriffe in der jeweiligen Stunde und die Größe des Inhalts der Seite. Um nun zu untersuchen, welche Stichwörter in einem Wörterbuch wie dem Wiktionary auffällig häufig nachgeschlagen werden, obwohl dies aus ihrer Häufigkeit im Korpus nicht zu erwarten ist, fassten wir zunächst die stündlichen Logfiledateien aus dem Jahr 2013 zu wöchentlichen Datensets zusammen, um sie besser handhabbar und die entsprechenden Analysen computational nicht zu ‚teuer‘ werden zu lassen. Um daraufhin die Wörter zu ermitteln, welche abseits ihrer Korpusfrequenz auffällig oft nachgeschlagen werden, haben wir ein Verfahren zur Trendbereinigung gewählt, welches es ermöglicht, solche ‚Ausreißer‘ in der Nachschlagefrequenz zu ermitteln.²⁵

3.3.2 Methode

Das Verfahren zu dieser Berechnung ist grob umrissen Folgendes: Zunächst wird für jedes Stichwort die durchschnittliche wöchentliche Suchhäufigkeit berechnet (durch ein Smoothing-Verfahren, das einen gleitenden Median berechnet, der z.B. aus dem Wert für den jeweiligen Tag und den Werten drei Tage davor und drei Tage danach besteht). Damit wird die Trendkomponente für dieses Stichwort berechnet. Die Idee ist nun, die Daten zu ‚detrenden‘, d.h. diese Trendkomponente aus den Daten herauszurechnen, um Auffälligkeiten in den Aufrufstatistiken auszuspielen. Dafür zieht man die Trendkomponente von der tatsächlichen Suchhäufigkeit ab (Suchhäufigkeit–Trendhäufigkeit). Um dann auch noch unterschiedliche Stichwörter (mit einer unterschiedlichen allgemeinen Suchhäufigkeit) ranken zu können, wird eine relative trendbereinigte Version gebildet: (Suchhäufigkeit–Trendhäufigkeit)/Trendhäufigkeit.

3.3.3 Ergebnis und Diskussion

Im Ergebnis zeigte sich, dass zum einen Stichwörter, die das Thema einer gesellschaftlichen lexikalisch-semantischen Diskussion sind, auffällig oft nachgeschlagen werden. Ein Stichwort, welches auf diese Weise auffiel, war z.B. das Stichwort „Furor“. Joachim Gauck hatte in einer Rede Anfang März von einem „Tugendfuror“

²³ S. <<http://wikimediafoundation.org>> [2014-02-21].

²⁴ Die URL des Index lautet <<http://dumps.wikimedia.org/other/pagecounts-raw/2013>> [2014-02-21].

²⁵ S. das entsprechende Arbeitspapier unter <http://www.ids-mannheim.de/fileadmin/lexik/pdf-download/Arbeitspapier_Trendbereinigung_Wiktionary-LF.pdf> [2014-03-15].

in Hinsicht auf die Debatte um Alltagssexismus gesprochen und damit eine Diskussion ausgelöst, ob dies eine angemessene Bezeichnung sei.²⁶ Dass so ein Wort daraufhin häufig nachgeschlagen wird, ist nicht besonders verwunderlich – ist es doch ein zumindest temporär gesellschaftlich sehr relevantes Wort. Einen entsprechenden Verlauf findet man auch ganz allgemein für die Google-Suche, wie sie Google in den „Google Trends“ veröffentlicht.²⁷ Ähnliches gilt für Wörter, die bei „Wer wird Millionär?“ diskutiert werden, wie am 6.5.2013 zur Wortherkunft von „Tribüne“, wozu es aufgrund des spannenden Verlaufs der Sendung und der Streitbarkeit der Lösung ein großes Medien-Echo gab.²⁸ (Temporäre) gesellschaftliche Relevanz, die sich in einer medialen Aufmerksamkeit spiegelt, scheint also ein weiterer Faktor zu sein, der einen Einfluss auf die Nachschlagehäufigkeit hat.

Erstaunlicher ist es, dass in dieser statistischen Analyse auch ein Stichwort wie ‚larmoyant‘ an einem Tag besonders auffällig wird. Die Spurensuche ergab, dass der Kommentator eines Spiels der deutschen Fußball-Nationalmannschaft (am 6.2.2013) bemerkte: „Der [Joachim Löw] ist jetzt aber richtig sauer. Das ist ihm ein bisschen zu larmoyant ...“.²⁹ Dies wiederum führt noch innerhalb derselben Stunde (und nur dieser) zu einem statistisch auffälligen Anstieg in der Abfrage dieses Stichworts (vgl. Abb. 9).

Dies erschien uns deshalb bemerkenswert, weil eine solche unmittelbare Verbindung zwischen dem Schauen eines Fußballspiels und Recherchieren im Wiktionary besteht – ein Zusammenhang, den es für gedruckte Wörterbücher so höchstwahrscheinlich nicht gegeben hat und der sicher auch mit der vermehrten Wörterbuchnutzung an mobilen Geräten in Zusammenhang steht. Genauso wurde das Wort ‚Borussia‘ immer häufiger nachgeschlagen, je weiter Borussia Dortmund in der Champions League kam (s. Abb. 10). Auch dies ist nicht unbedingt zu erwarten, weil das Wort ‚Borussia‘ kein Gegenstand von Bedeutungsdiskussion im engeren Sinne ist. Solche Beobachtungen weiterzuführen, ist sicher eine spannende Forschungsaufgabe auch für die Zukunft.

26 S. z. B. „Zeit Online“: (<<http://www.zeit.de/gesellschaft/2013-03/aufschrei-gauck-tugendfuror>>); „Süddeutsche Online“: (<<http://www.sueddeutsche.de/politik/sexismus-debatte-als-tugendfuror-aufschrei-wegen-gauck-1.1616310>>) oder „Welt Online“: (<<http://www.welt.de/politik/deutschland/article114192084/Sexismus-Vorwurf-wegen-Gauks-Tugendfuror.html>>) [2014-02-21].

27 S. für „Furor“ und „Tugendfuror“: <<http://www.google.com/trends/explore?hl=en-US#q=Furor%2C%20Tugendfuror&geo=DE&cmpt=q>> [2014-02-21].

28 S. z. B. <<http://www.jurpc.de/jurpc/show?id=20130087>> oder <http://www.focus.de/kultur/kino_tv/tid-31164/verwirrung-um-jauch-show-rtl-wehrt-sich-gegen-wer-wird-millionaer-desaster_aid_987893.html> [2014-02-21].

29 Leider können hierzu keine frei zugänglichen Nachweise gegeben werden. Die Sportschau-Redaktion war so freundlich, uns einen DVD-Mitschnitt dieses Spiels zuzusenden.

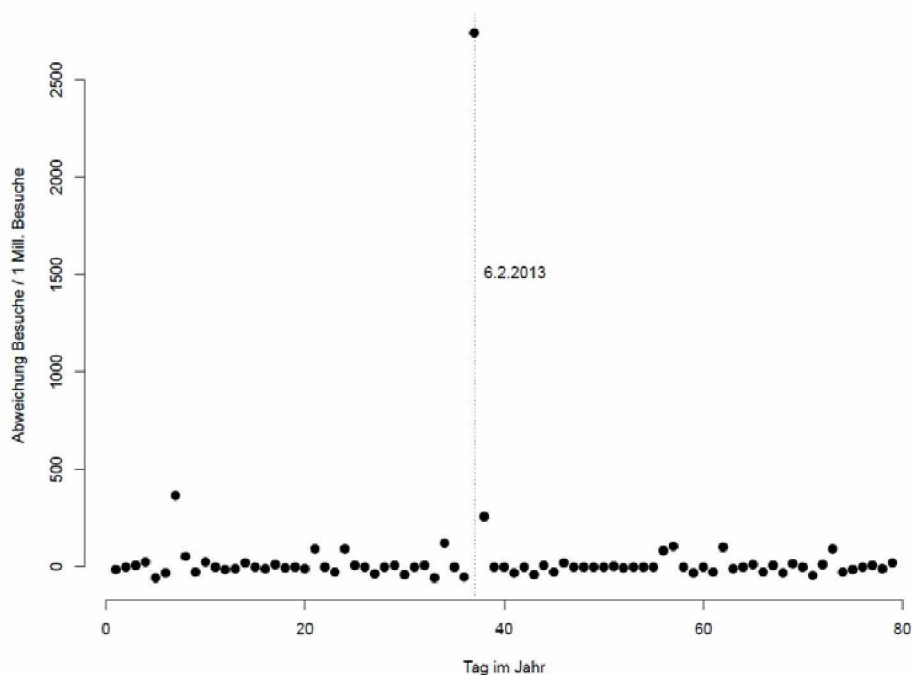


Abb. 9: Trendbereinigte Zugriffe für „larmoyant“ (1-12/2013)

Eine solche Trenderhebung basierend auf Logfiles kann also interessante Ergebnisse über das Nachschlageverhalten geben, insbesondere wenn es darum geht, welcher Teil des Wortschatzes besonders häufig nachgeschlagen wird. Wenig lässt sich dagegen dazu sagen, welche Wortartikel besonders intensiv rezipiert werden, welche äußeren Benutzungskontexte zu der Wörterbuchbenutzung geführt haben etc. Deshalb wird von manchen Autorinnen empfohlen, die Analyse von Logfiles mit anderen Möglichkeiten der Datenerhebung zu kombinieren, z.B. die Benutzer, die ein Wörterbuch benutzen, nachher zu bitten, einen Fragebogen zu beantworten, und dieses Antwortverhalten mit den Logfileinformationen zu kombinieren (Hult 2012; Tarp 2009, 290). Allerdings ist ein solches Vorgehen – zumindest in Deutschland – aus datenschutzrechtlichen Beschränkungen her nicht möglich, es sei denn, man weist die Benutzer vor dem Aufrufen von Wörterbuchinhalten darauf hin, dass alle ihre Zugriffe auf Inhalte personalisiert protokolliert werden, was aber wiederum ihr Nachschlageverhalten beeinflussen könnte. Ansonsten müssen bei Logfileanalysen personalisierte Informationen wie IP-Adressen anonymisiert werden, sodass eine nachträgliche personalisierte Analyse des Nachschlageverhaltens einer bestimmten Nutzerin, der auch einen Fragebogen beantwortet hat, nicht möglich ist.

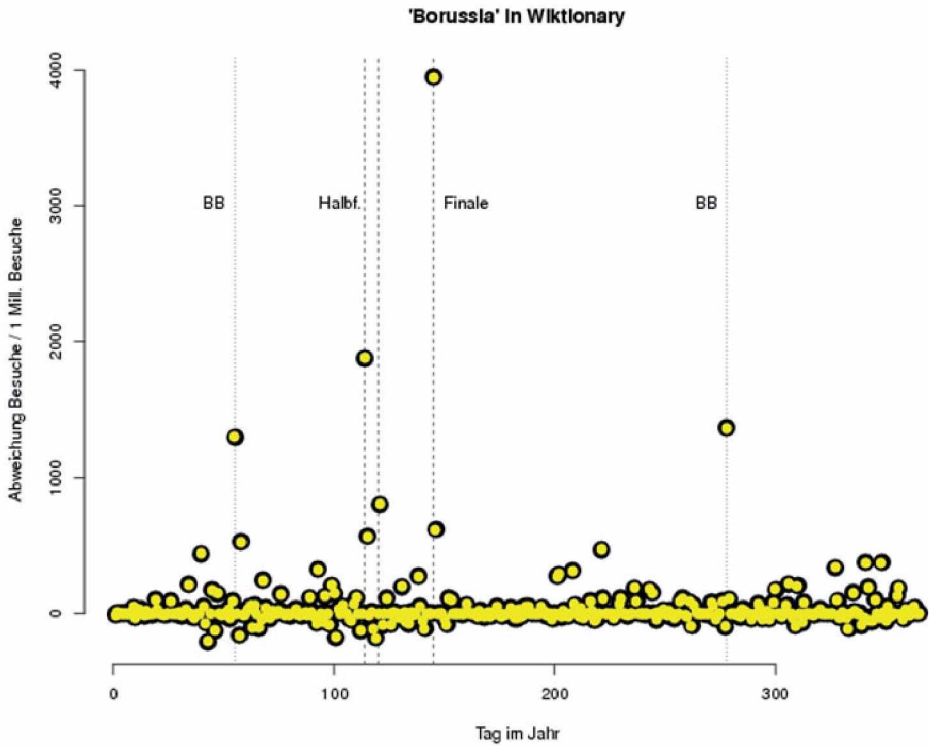


Abb. 10: Trendbereinigte Zugriffe für „Borussia“ (1-12/2013). Vertikale Linien kennzeichnen Fußballspiele (BB: Borussia Mönchengladbach vs. Borussia Dortmund, die anderen die Spiele der Champions League).

3.4 Wie navigieren potenzielle Nutzer zu einzelnen Bedeutungen im Onlinewörterbuch *ellexiko*?

In diesem Abschnitt wird anhand eines Studienausschnitts eine weitere, relativ neue Form der Beobachtung vorgestellt, und zwar die Datenerhebung in Form von Eyetracking. Mit Eyetracking bezeichnet man das Aufzeichnen der hauptsächlich aus Fixationen (Punkte, die man genau betrachtet), Sakkaden (schnellen Augenbewegungen) und Regressionen bestehenden Blickbewegungen einer Person; die entsprechenden Geräte, die diese Aufzeichnungen vornehmen, nennt man Eyetracker. Abbildung 11 zeigt ein Werbefoto eines Eyetrackers, wie wir ihn in unserer Studie eingesetzt haben. Eyetracking ist eine reaktive Methode, da die Probandin weiß, dass ihr Verhalten aufgezeichnet wird. Was Eyetracking als Datenerhebungsmethode von anderen reaktiven Verfahren abhebt, ist, dass das Verhalten in einer Eyetrackingstudie unmittelbar aufgezeichnet wird. Das Verfahren erlaubt damit Rückschlüsse über Vorgänge bei den

Versuchspersonen, die noch unmittelbarer sind als etwa bei der Methode des ‚lauten Denkens‘. Dadurch, dass darüber hinaus Augenbewegungen schwer zu kontrollieren sind, können Effekte wie etwa soziale Erwünschtheit oder der Versuchsleitereffekt, die bei Nutzerbefragungen und Fragebögen auftreten können, beim Eyetracking ausgeschlossen, bzw. minimiert werden.

Das Ziel unserer Eyetracking-Studie war zum einen, im Rahmen unseres Forschungsprojekts zur Wörterbuchbenutzungsforschung diese Methode der Datenerhebung auszuprobieren und Erfahrungen in diesem Bereich zu sammeln. Zum anderen wollten wir den Neuauftritt des Wörterbuchportals OWID,³⁰ den wir intern fertiggestellt, aber zu der Zeit noch nicht online freigeschaltet hatten, evaluieren.

Um eine Eyetracking-Studie durchführen zu können, braucht man zunächst ein entsprechend ausgestattetes Labor. Am Institut für Deutsche Sprache haben wir kein solches, deshalb führten wir die Studie in Kooperation mit der Universität Mannheim (Lehrstuhl Rosemarie Tracy) durch.³¹ Das Labor dort verfügt über verschiedene Computerstationen mit einem Eyetracker, der für Lesezeitexperimente geeignet ist (für diese braucht man eine sehr genaue Auflösung, da man z.B. zeilen- und wortgenau sehen muss, welche Teile eines Textes genau gelesen werden), und einem sogenannten *Remote Eye Tracker* der Marke SMI RED, der für allgemeine Blickbeobachtungsstudien geeignet ist und den wir auch für unsere Studie benutzten. Der kleine Kasten unter dem Bildschirm dient dabei dem Aufzeichnen der Blickbewegungen. Der jeweilige Proband saß dabei jeweils vor dem Eyetracker; die Versuchsleiterin saß durch eine Trennwand abgetrennt im gleichen Raum daneben. Sie musste während des Versuchs kontrollieren, dass die Teilnehmer sich nicht aus dem ‚Sichtfeld‘ des Eyetrackers hinausbewegten. Das Setting, d.h. der Versuchsaufbau, war für die Probandinnen daher verhältnismäßig natürlich, da keine auffälligen Gerätschaften eingesetzt werden mussten, wie es auch bei Lew für eine Eyetracking-Studie berichtet wurde:

Thanks to these features, the Tobii T60 has high ecological validity, offering participants the look and feel of a regular computer screen, thus a highly naturalistic setting for students accustomed to working at the computer. (Lew et al. 2013, 11)

Dagegen berichtet beispielsweise Tono:

While the eye mark recorder is a powerful tool, the setting inevitably becomes artificial. In order to calculate gaze points accurately, it was necessary to fix the subjects' head onto the chinrest and ask them to look at the PC monitor. (Tono 2011, 151)

³⁰ <www.owid.de> [2014-02-21].

³¹ Siehe <http://master.phil.uni-mannheim.de/masterstudiengaenge/master_sprache_und_kommunikation/experimentallabor_und_mai_lab/index.html> [2014-02-21].

An unserer Studie nahmen 38 Personen im Alter von 20–30 Jahren teil. Alle Teilnehmer erhielten 10 Euro für ihre Teilnahme. Durchgeführt wurde die Studie im August/September 2011. Fast 40 Teilnehmer sind sehr viele für eine Eyetracking-Studie; andere Eyetracking-Studien aus dem Bereich der Wörterbuchbenutzungsforschung hatten 6–8 Probanden (z.B. Simonsen 2011; Tono 2011).



Abb. 11: Werbefoto des SMI Eyetrackers (<<http://www.gizmag.com/smi-red500-500hz-remote-eye-tracker/16957/picture/124519/>> [2014-02-21].)

Ein Bereich, den wir in unserer Eyetracking-Studie untersuchen wollten, waren bestimmte Elemente der inneren Zugriffsstruktur, die wir im neuen Webdesign verändert hatten. Eine davon bezog sich auf die Navigation zu den Einzelbedeutungen in *elexiko*, einem der Wörterbücher aus OWID. Diese Forschungsfrage und die zugehörigen Ergebnisse sollen hier dargestellt werden.

In *elexiko* sind die Angaben zu einem Stichwort in einer Art Reiteransicht auf verschiedene Bildschirme verteilt: auf der ersten Seite finden sich einzelbedeutungsübergreifende Informationen wie die Schreibung eines Wortes, Angaben zur Worttrennung, zur Wortbildung etc. Die Informationen zu den einzelnen Bedeutungen wie typische Verwendungen, sinnverwandte Wörter etc. folgen auf einem zweiten Bildschirm, wenn man auf die Etiketten der Einzelbedeutungen klickt (vgl. Abb. 12). Im alten OWID-Layout waren dabei auf dem ersten Bildschirm nur Kurzetiketten zu den Einzelbedeutungen gezeigt. Im neuen Layout wurde dies dahingehend geändert, dass wir zusätzlich zu den Etiketten schon auf dem ersten Bildschirm die Paraphrase platzierten. Dies sollte den Benutzern helfen, einen schnelleren Eindruck über das Bedeutungsspektrum und die für sie in der jeweiligen Benutzungssituation relevante Einzelbedeutung zu bekommen.



Abb. 12: Lesartenübergreifende (links) und lesartenbezogene Angaben (rechts) in elexiko

In der Eyetrackingstudie wollten wir untersuchen, wie die Teilnehmer diese Informationen wahrnehmen, genauer formuliert: Wie sehen die Blickbewegungsmuster aus, wenn wir den Probandinnen Fragen zu Einzelbedeutungen stellen? Finden sie die relevanten Einzelbedeutungen? Lesen bzw. scannen sie alle Etikettierungen zuerst und lesen dann erst die Paraphrasen? Oder ist es (obwohl das sehr unwahrscheinlich ist) ein linearer Leseprozess? Unser Wunsch bei der Entwicklung des neuen Designs dieser Angaben war es, dass die Etikettierungen zuerst ‚ins Auge springen‘, und erst bei Bedarf die gesamte Paraphrase gelesen wird. Wenn sich dies in den Blickverläufen der Teilnehmer bestätigen würde, konnten wir dies demnach als Bestätigung unseres Designs auffassen.³²

Die Vorgehensweise für die Untersuchung war folgendermaßen: In einer ersten Aufgabe baten wir die Teilnehmerinnen zu schauen, ob das Stichwort „Pferd“ eine Einzelbedeutung ‚Turngerät‘ hat: „*Sie sehen auf der nächsten Seite einen Wortartikel aus elexiko. Bitte versuchen Sie herauszufinden, ob das Stichwort eine Bedeutung/Lesart im Sinne von ‚Turngerät‘ hat.*“ Damit wollten wir schauen, ob die Probanden die

³² Zum gezielten Suchen von Einzelbedeutungen, vor allem in gedruckten Wörterbüchern, gibt es bereits Studien in der Wörterbuchbenutzungsforschung (Lew/Tokarek 2010; Lew/Grzelak/Leszakowicz 2013, für eine Zusammenfassung anderer Studien aus dem Bereich s. bes. S. 4–6; Lew 2010; Nesi/Tan 2011; Tono 2001; Tono 2011b). Allerdings war unser Studiendesign anders als in den genannten Studien, so dass die Ergebnisse schlecht vergleichbar sind.

relevante Einzelbedeutung schnell auffinden. Das Ergebnis ist in Abb. 13 zu sehen. Man sieht auf der linken Seite eine sogenannte Heatmap, die die totale Dauer der Betrachtung eines Bereiches, kumuliert von allen Teilnehmern anzeigt; die Fixationsdauer wird dabei durch eine entsprechende Einfärbung verdeutlicht. Man sieht hier deutlich, dass die Aufmerksamkeit auf die relevante Einzelbedeutung konzentriert war. Auf der rechten Seite der Abbildung 13 ist ein ‚Scan path‘ einer Probandin zu sehen. Hier kann man erkennen, in welchen Fixationsschritten dieser Proband gesucht hat. Insgesamt kann man aus den Eyetracking-Daten ablesen, dass die relevante Einzelbedeutung schnell gefunden wurde. Allerdings war dies auch eine einfache Aufgabe.

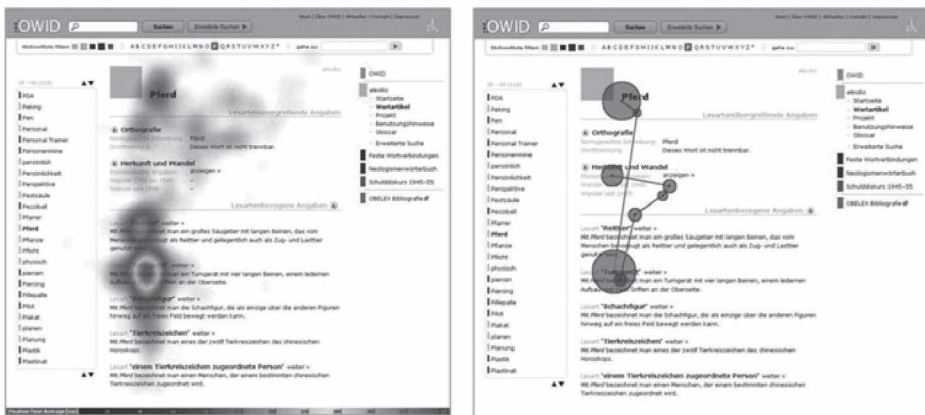


Abb. 13: Heat map aller Probanden (links); Scan Pfad eines Probanden (rechts) (Aufgabe: Auffinden der Einzelbedeutung ‘Turngerät’)

In einem zweiten Schritt baten wir die Teilnehmerinnen, eine bestimmte Einzelbedeutung des Stichworts „Mannschaft“ zu finden: *„Bitte versuchen Sie herauszufinden, ob es im folgenden Wortartikel eine Bedeutung gibt, die erläutert ist mit ‚Mitglieder einer für eine Organisation tätige Gruppe von Menschen‘. Wenn ja, welche?“* Das Ergebnis ist in Abb. 14 gezeigt.

Das Interessante ist hier, dass die Teilnehmerinnen offensichtlich zuerst die Etiketten sehr schnell scannen (beide hier gezeigten Teilnehmer hatten bereits nach 00:01 Sekunden alle Etiketten gescannt), und dann erst die Aufmerksamkeit auf die Paraphrase lenken. Dies entspricht der Vorgehensweise, wie wir sie uns gewünscht haben für das neue Design. Insgesamt kann man als Ergebnis zu diesem kleinen Ausschnitt der Eyetracking-Studie also festhalten, dass die Probandinnen die relevanten Einzelbedeutungen gefunden haben und die unterschiedlichen Funktionen von Etiketten und Paraphrasen, wie sie im neuen Design angedacht wurden, in der Praxis deutlich wurden.

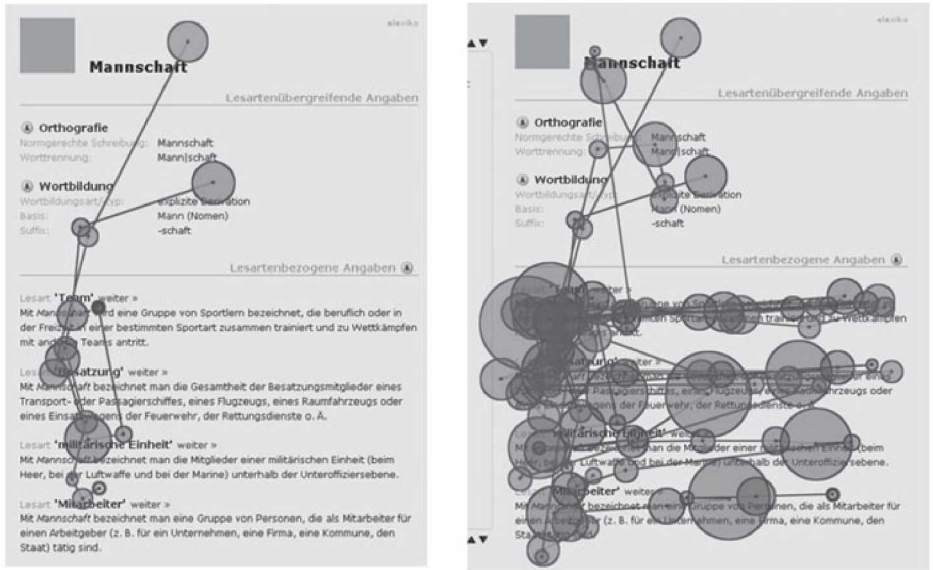


Abb. 14: Scan paths von zwei Teilnehmern (gespeichert als Film); ein Schnappschuss bei 00:01 Sekunden (links) und der zweite bei 00:14 Sekunden (rechts)

Eine Anmerkung noch am Rande: In unserem Team hatte keiner Erfahrung mit dieser Form der Datenerhebung; diese mussten wir erst sammeln. Wir haben zum Beispiel erst bei der Analyse gemerkt, dass wir besser mehr vergleichende Ansichten von dem alten gegenüber dem neuen Layout eingesetzt hätten, um wirklich schließen zu können, dass das neue Layout besser funktioniert als das alte. So wie wir es gemacht haben, konnten wir oftmals nur schlussfolgern, dass das neue Layout – wie im eben gezeigten Fall – gut funktioniert, aber vielleicht hätte das alte Layout dies genauso getan?³³ Aber auch solche Lernprozesse gehören zur Forschungsarbeit.

4 Schlussbemerkung

Die vorangegangenen Beispiele für empirische Studien aus dem Bereich der Wörterbuchbenutzungsforschung konnten hoffentlich verdeutlichen, dass es eine große Bandbreite an Fragestellungen und dazu passenden Untersuchungsmethoden in diesem Forschungsfeld gibt. Aktuell ist die Erforschung der Benutzung von Wörterbüchern auch deshalb besonders wichtig, da die Lexikografie sich in einer exis-

³³ Diese methodischen Mängel werden in Müller-Spitzer/Michaelis/Koplenig 2014 ausführlich dargestellt.

tentiellen Krise befindet. In Zeiten kostenloser Onlinewörterbücher werden immer weniger Wörterbücher gekauft, sodass die Verlage große Schwierigkeiten haben, ihr Personal und ihre Ressourcen weiter zu halten. Auch die öffentliche Hand finanziert kaum noch über Jahrzehnte andauernde lexikografische Großprojekte.

Gleichzeitig werden sehr viele sprachliche Fragen im Internet recherchiert – vielleicht oder sehr wahrscheinlich mehr – als es Nachschlagehandlungen in gedruckten Wörterbüchern gab. Die Frage ist also, wie man diese Nutzungshandlungen besser mit den vorhandenen lexikografischen Ressourcen zusammenbringen kann. Neben der Untersuchung aktueller Wörterbücher ist daher für die Zukunft innerhalb der Benutzungsforschung besonders wichtig, den Ausgangspunkt der Wörterbuchbenutzung nicht aus dem Auge verlieren, d.h. die Situationen, in denen sprachliche Schwierigkeiten aufkommen und aus denen heraus dann ein Nachschlagebedürfnis entsteht. Denn im Grunde muss man sich, wenn man mit der Benutzungsforschung dafür sorgen möchte, dass Wörterbücher besser den tatsächlichen Nutzungsbedürfnissen entsprechen, genau bei diesen Nutzungsbedürfnissen anfangen.

Theodore Levitt, ein US-amerikanischer Ökonom, schrieb in den 60er-Jahren einen einflussreichen Artikel mit dem Titel „Marketing Myopia“, in dem er genau auf diesen Aspekt hinweist: nämlich, dass es auch in der Industrie darum geht, sich nicht auf ein Produkt bzw. eine Produktart einzugrenzen, sondern sich auf den Zweck zu konzentrieren, für den dieses Produkt entwickelt wurde:

The railroads did not stop growing because the need for passenger and freight transportation declined. That grew. The railroads are in trouble today not because the need was filled by others (cars, trucks, airplanes, even telephones), but because it was not filled by the railroads themselves. They let others take customers away from them because they assumed themselves to be in the railroad business rather than in the transportation business. The reason they defined their industry wrong was because they were railroad-oriented instead of transportation-oriented; they were product-oriented instead of customer-oriented. (Levitt 1960, 24)

Bezogen auf die Wörterbuchbenutzungsforschung könnte dies heißen, dass sie ihren Blick erweitern sollte: über die Untersuchung der Benutzung heute vorliegender Wörterbücher hinaus hin zu den sprachlichen Schwierigkeiten, in denen die Nachschlagebedürfnisse entstehen. Dabei bin ich wie auch Lew davon überzeugt, dass solche Untersuchungen nicht allein mit der Methode der ‚Deduktion‘, d.h. mit der Befragung von Experten zu bewältigen sind.

The studies [...] here show over and over again that expert opinion, intuition, or purely deductive reasoning cannot replace solid empirical evidence from user studies: dictionary use is just too complex an affair to be that predictable. (Lew 2011a, 3)

5 Bibliographie

- Bank 2012 = Bank, Christina: Die Usability von Online-Wörterbüchern und elektronischen Sprachportalen. In: *Information, Wissenschaft & Praxis* 63/6. 2012, 345–360.
- Bergenholtz 2011 = Bergenholtz, Henning: Access to and Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries. In: Bergenholtz, Henning / Fuertes-Olivera, Pedro Antonio (eds.): *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London / New York: Continuum, 2011, 30–53.
- Bergenholtz/Bergenholtz 2011 = Bergenholtz, Henning / Bergenholtz, Inger: A Dictionary Is a Tool, a Good Dictionary Is a Monofunctional Tool. In: Bergenholtz, Henning / Fuertes-Olivera, Pedro Antonio (eds.): *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London / New York: Continuum, 2011, 187–207.
- Bergenholtz/Bothma 2011 = Bergenholtz, Henning / Bothma, Theo J.D.: Needs-adapted data presentation in e-information tools. In: *Lexikos* 21. 2011, 53–77.
- Bergenholtz/Johnsen 2005 = Bergenholtz, Henning / Johnsen, Mia: Log Files as a Tool for Improving Internet Dictionaries. In: *Hermes. Journal of Language and Communication Studies* 34. 2005, 117–141.
- Bergenholtz/Tarp 2003 = Bergenholtz, Henning / Tarp, Sven: Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. In: *Hermes. Journal of Language and Communication Studies* 31. 2003, 171–196.
- Bogaards 2003 = Bogaards, Paul: Uses and users of dictionaries. In: van Sterkenburg, Piet (ed.): *A Practical Guide to Lexicography*. Amsterdam / Philadelphia: John Benjamins Publishing Company, 2003, 26–33.
- Bowker 2012 = Bowker, Lynne: Meeting the needs of translators in the age of e-lexicography: Exploring the possibilities. In: Granger, Sylviane / Paquot, Magali (eds.): *Electronic lexicography*. Oxford: Oxford University Press, 2012, 379–397.
- De Schryver 2003 = De Schryver, Gilles-Maurice: Lexicographers' Dreams in the Electronic Dictionary Age. In: *International Journal of Lexicography* 16/2. 2003, 143–199.
- De Schryver/Joffe 2004 = De Schryver, Gilles-Maurice / Joffe, David: On how electronic dictionaries are really used. In: Williams, Geoffrey / Vessier, Sandra (eds.): *Proceedings of the XI EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10. Lorient: UBS, 2004, 187–196.*
- De Schryver et al. 2006 = De Schryver, Gilles-Maurice / Joffe, David / Joffe, Pitta / Hillewaert, Sarah: Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography. In: *Lexikos* 16. 2006, 67–83.
- Diekmann 2011 = Diekmann, Andreas: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen* 5. Aufl. Hamburg: Rowohlt, 2011.
- Dziemianko 2012 = Dziemianko, Anna: On the use(fulness) of paper and electronic dictionaries. In: Granger, Sylviane / Paquot, Magali (eds.): *Electronic lexicography*. Oxford: Oxford University Press, 2012, 320–341.
- Heid/Zimmermann 2012 = Heid, Ulrich / Zimmermann, Jan Timo: Usability testing as a tool for e-dictionary design: collocations as a case in point. In: Torjusen, Julie Matilde / Fjeld, Ruth Vatvedt (eds.): *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo. Oslo: UiO, 2012, 661–671.*
- Householder 1962 = Householder, Fred W.: *Problems in Lexicography*. Bloomington: Indiana University Press, 1962.
- Hult 2012 = Hult, Ann-Kristin: Old and New User Study Methods Combined – Linking Web Questionnaires with Log Files from the Swedish Lexin Dictionary. In: Fjeld, Ruth Vatvedt / Torjusen, Julie Matilde (eds.): *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo. Oslo: UiO, 2012, 922–928.*

- Klosa/Koplenig/Töpel 2014 = Klosa, Annette / Koplenig, Alexander/Töpel, Antje: Benutzerwünsche und -meinungen zu dem monolingualen deutschen Onlinewörterbuch *elexiko*. In: Müller-Spitzer, Carolin (ed.): *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014, 281–384.
- Koplenig 2014 = Koplenig, Alexander: Empirical research into dictionary use. In: Müller-Spitzer, Carolin (ed.): *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014, 55–76.
- Koplenig/Meyer/Müller-Spitzer 2014 = Koplenig, Alexander / Meyer, Peter / Müller-Spitzer, Carolin: Dictionary users do look up frequent words. A log file analysis. In: Müller-Spitzer, Carolin (ed.): *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014, 229–250.
- Koplenig/Müller-Spitzer 2014 = Koplenig, Alexander / Müller-Spitzer, Carolin: Questions of design. In: Müller-Spitzer, Carolin (ed.): *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014, 189–204.
- Levitt 1960 = Levitt, Theodore: Marketing Myopia. In: *Harvard Business Review* 38. 1960, 24–47.
- Lew 2011a = Lew, Robert: Studies in Dictionary Use: Recent Developments. In: *International Journal of Lexicography* 24/1, 2011, 1–4.
- Lew 2011b = Lew, Robert: User studies: Opportunities and limitations. In: Akasu, Kaoru / Satoru, Uchida (eds.): *ASIALEX 2011 Proceedings Lexicography: Theoretical and practical perspectives*. Kyoto: Asian Association for Lexicography, 2011, 7–16.
- Lew 2010 = Lew, Robert: Users Take Shortcuts: Navigating Dictionary Entries. In: Dykstra, Anne / Schoonheim, Tanneke (eds.): *Proceedings of the XIV Euralex International Congress* (Leeuwarden, 6–10 July 2010). Ljouwert: Fryske Akademy – Afûk, 2010, 1121–1132.
- Lew/Grzelak/Leszkowicz 2013 = Lew, Robert / Grzelak, Marcin / Leszkowicz, Mateusz: How Dictionary Users Choose Senses in Bilingual Dictionary Entries: An Eye-Tracking Study. In: *Lexikos* 23. 2013, 228–254.
- Lew/Tokarek 2010 = Lew, Robert / Tokarek, Patryk: Entry menus in bilingual electronic dictionaries. In: *eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve: Cahiers du Cental, 2010, 193–202.
- Mayring 2011 = Mayring, Philipp: *Qualitative Inhaltsanalyse. Grundlagen und Techniken* 8. Aufl.. Weinheim: Beltz, 2011.
- Müller-Spitzer 2013 = Müller-Spitzer, Carolin: Contexts of dictionary use. In: Kosem, Iztok / Kallas, Jelena / Gantar, Polona / Krek, Simon (eds.): *Electronic lexicography in the 21st century: thinking outside the paper*. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia. Ljubljana / Tallinn: Trojina, Institute for Applied Slovene Studies / Eesti Keele Instituut, 2013.
- Müller-Spitzer 2014a = Müller-Spitzer, Carolin: *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014.
- Müller-Spitzer 2014b = Müller-Spitzer, Carolin: Empirical data on contexts of dictionary use. In: Müller-Spitzer, Carolin (ed.): *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014, 85–126.
- Müller-Spitzer/Koplenig 2014 = Müller-Spitzer, Carolin / Koplenig, Alexander (2014): Online dictionaries: expectations and demands. In: Müller-Spitzer, Carolin (ed.): *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014, 143–188.
- Müller-Spitzer/Michaelis/Koplenig 2014 = Müller-Spitzer, Carolin / Michaelis, Frank / Koplenig, Alexander: Evaluation of a new web design for the dictionary portal OWID An attempt at using eye-tracking technology. In: Müller-Spitzer, Carolin (ed.): *Using Online Dictionaries*. Berlin / Boston: de Gruyter, 2014, 207–228.
- Nesi/Tan 2011 = Nesi, Hilary / Tan, Kim Hua: The Effect Of Menus And Signposting On The Speed And Accuracy Of Sense Selection. In: *International Journal of Lexicography* 24/1. 2011, 79–96.
- Nesi 2012 = Nesi, Hilary: Alternative e-dictionaries: Uncovering dark practices. In: Granger, Sylviane/Paquot, Magali (eds.): *Electronic lexicography*. Oxford: Oxford University Press, 2012, 363–378.

- Prinsloo et al. 2012 = Prinsloo, Danie J. / Heid, Ulrich / Bothma, Theo / Faaß, Gertrud: Devices for Information Presentation in Electronic Dictionaries. In: *Lexikos* 22. 2012, 290–320.
- Rundell 2012a = Rundell, Michael: It works in practice but will it work in theory? The uneasy relationship between lexicography and matters theoretical. In: Fjeld, Ruth Vatvedt / Torjusén, Julie Matilde (eds.): *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo*. Oslo: UiO, 2012, 47–92.
- Rundell 2012b = Rundell, Michael: The road to automated lexicography: An editor's viewpoint. In: Granger, Sylviane/Paquot, Magali (eds.): *Electronic lexicography*. Oxford: Oxford University Press, 2012, 15–30.
- Sharifi 2012 = Sharifi, Saghar: General Monolingual Persian Dictionaries and Their Users: A Case Study. In: Fjeld, Ruth Vatvedt / Torjusén, Julie Matilde (eds.): *Proceedings of the 15th EURALEX International Congress, 7–11 August, 2012, Oslo*. Oslo: UiO, 2012, 626–639.
- Simonsen 2011 = Simonsen, Henrik K  hler: User Consultation Behaviour in Internet Dictionaries: An Eye-Tracking Study. In: *Hermes. Journal of Language and Communication Studies* 46. 2011, 75–101.
- Sollaci/Pereira 2004 = Sollaci, Luciana B. / Pereira, Mauricio G. (2004): The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. In: *Journal of the Medical Library Association* 92/3, 2004, 364–371.
- Tarp 2008 = Tarp, Sven: *Lexicography in the borderland between knowledge and non-knowledge: general lexicographical theory with particular focus on learner's lexicography*. T  bingen: Max Niemeyer Verlag, 2008.
- Tarp 2009 = Tarp, Sven: Reflections on Lexicographical User Research. In: *Lexikos* 19. 2009, 275–296.
- Tarp 2011 = Tarp, Sven: Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. In: Bergenholtz, Henning / Fuertes-Olivera, Pedro Antonio (eds.): *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London / New York: Continuum, 2011, 54–70.
- Tono 2000 = Tono, Yukio: On the effects of different types of electronic dictionary interfaces on L2 learners' reference behaviour in productive/receptive tasks. In: Heid, Ulrich / Evert, Stefan / Lehmann, Egbert / Rohrer, Christian (eds.): *Proceedings of the IX EURALEX International Congress, EURALEX 2000*. Stuttgart, Germany, August 8th–12th, 2000. Stuttgart: IMS, 2000, 855–861.
- Tono 2001 = Tono, Yukio: Research on dictionary use in the context of foreign language learning: Focus on reading comprehension. T  bingen: Max Niemeyer Verlag, 2001.
- Tono 2011 = Tono, Yukio: Application of Eye-Tracking in EFL Learners' Dictionary Look-up Process Research. In: *International Journal of Lexicography* 24/1, 2011, 124–153.
- Trap-Jensen 2010 = Trap-Jensen, Lars: One, Two, Many: Customization and User Profiles in Internet Dictionaries. In: Dykstra, Anne / Schoonheim, Tanneke (eds.): *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*. Ljouwert: Fryske Akademy – Af  k, 2010, 1133–1143.
- Trap-Jensen/Lorentzen 2011 = Trap-Jensen, Lars/Lorentzen, Henrik: There And Back Again – from Dictionary to Wordnet to Thesaurus and Vice Versa: How to Use and Reuse Dictionary Data in a Conceptual Dictionary. In: Kosem, Iztok / Kosem, Karmen (eds.): *Electronic lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex2011*. Bled / Ljubljana 2011, 175–179.
- Verlinde/Peeters 2012 = Verlinde, Serge / Peeters, Geert: Data access revisited: The Interactive Language Toolbox. In: Granger, Sylviane / Paquot, Magali (eds.): *Electronic lexicography*. Oxford: Oxford University Press, 2012, 147–162.
- Welker 2010 = Welker, Herbert Andreas: *Dictionary use: a general survey of empirical studies*. Bras  lia: Eigenverlag, 2010.

- Wiegand 1977 = Wiegand, Herbert Ernst: Nachdenken über Wörterbücher: Aktuelle Probleme. In: Drosdowski, Helmut / Henne, Helmut / Wiegand, Herbert: Nachdenken über Wörterbücher, Mannheim: Bibliographisches Institut, 1977, 51–102.
- Wiegand 1998 = Wiegand, Herbert Ernst: Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. Berlin / New York: de Gruyter, 1998.
- Wiegand et al. 2010 = Wiegand, Herbert Ernst / Beißwenger, Michael / Gouws, Rufus H. / Kammerer, Matthias et al.: Wörterbuch zur Lexikographie und Wörterbuchforschung. Mit englischen Übersetzungen der Umtexpte und Definitionen sowie Äquivalenten in neuen Sprachen [...]. Berlin / New York: de Gruyter, 2010.
- Wolfer/Koplenig/Müller-Spitzer (erscheint) = Wolfer, Sascha / Koplenig, Alexander / Meyer, Peter / Müller-Spitzer, Carolin (2014): Dictionary users do look up frequent and socially relevant words. Two log-file analyses. In: Abel, A. / Vettori, C. / Ralli, N. (Hrsg.): Proceedings of the XVI EURALEX International Congress: The User in Focus. Bolzano/Bozen, 281–290.